DOCUMENT RESUME

ED 146 232 95 TM 006 638

AUTHOR Nesselroade, John R.; Baltes, Paul B.
TITLE Longitudinal Research in the Behavioral Sciences: Design and Analysis. Final Report.
INSTITUTION Pennsylvania State Univ., University Park.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Basic Skills Group.
BUREAU NO BR-4-1115
PUB DATE 1 Jun 77
CONTRACT NIE-C-74-0127
NOTE 468p.

EDRS PRICE MF-$0.83 HC-$24.77 Plus Postage.
DESCRIPTORS Behavioral Science Research; *Data Analysis; *Educational Research; *Human Development; *Longitudinal Studies; *Mathematical Models; Research Design; Research Methodology; Research Problems; *Statistical Analysis

ABSTRACT

This manual is intended to improve both the design of longitudinal studies and analysis of the resulting data. Issues related to educational and developmental research have been emphasized in these eight chapters. Topics of particular interest to longitudinal researchers include stochastic models of developmental change, mathematical representation of growth curves, structural equation models, analysis of univariate and multivariate variance, and analysis of categorical data. The differentiation between intraindividual variation, interindividual differences, and interindividual differences in intraindividual change is also discussed. (Author/GDC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED146232

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

PROJECT NO. 4-1115
CONTRACT NO. NIE-C-74-0127

LONGITUDINAL RESEARCH IN THE BEHAVIORAL SCIENCES: DESIGN AND ANALYSIS

JOHN R. NESSELROADE AND PAUL B. BALTES
THE PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802

JUNE 1, 1977

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. DEPARTMENT OF

HEALTH, EDUCATION AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

BASIC SKILLS

TM006 638

FINAL REPORT

PROJECT NO. 4-1115
CONTRACT NO. NIE-C-74-0127

LONGITUDINAL RESEARCH IN THE BEHAVIORAL SCIENCES: DESIGN AND ANALYSIS

JOHN R. NESSELROADE AND PAUL B. BALTES
THE PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802

JUNE 1, 1977

The research reported herein was performed pursuant to a grant with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U.S. DEPARTMENT OF

HEALTH, EDUCATION AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

BASIC SKILLS

TABLE OF CONTENTS

PAGE

ABSTRACT i

PREFACE ii

COMPONENTS OF DEVELOPMENTAL CHANGE: INTRAINDIVIDUAL CHANGE AND INTERINDIVIDUAL DIFFERENCES. 1
ALLAN R. BUSS

MATHEMATICAL REPRESENTATIONS OF DEVELOPMENT THEORIES 35
BURTON SINGER and SEYMOUR SPILERMAN

MATHEMATICAL DESCRIPTION AND REPRESENTATION OF DEVELOPMENTAL CHANGE FUNCTIONS ON THE INTRA- AND INTERINDIVIDUAL LEVELS. 77
KENNETH E. GUIRE and CHARLES J. KOWALSKI

ASSESSMENT AND STATISTICAL CONTROL OF SUBJECT VARIABLES IN LONGITUDINAL DESIGNS 113
PAUL A. GAMES

UNIVARIATE AND MULTIVARIATE ANALYSIS OF VARIANCE OF TIME-STRUCTURED DATA. 149
R. DARRELL BOCK

THE ANALYSIS OF CATEGORICAL DATA IN LONGITUDINAL STUDIES OF BEHAVIORAL DEVELOPMENT 205
J. RICHARD LANDIS and GARY G. KOCH

CAUSAL MODELS IN LONGITUDINAL RESEARCH 261
DAVID ROGOSA

STATISTICAL ESTIMATION OF STRUCTURAL MODELS IN LONGITUDINAL-DEVELOPMENTAL INVESTIGATION. 308
KARL G. JÖRESKOG

REFERENCES 411

ABSTRACT

The focus of this project has been the preparation of a manual aimed at the improvement of both the design of longitudinal research and the analysis of data arising therefrom. Primary substantive emphasis has been given to issues related to educational and developmental research. Topics of particular salience to longitudinal researchers include: (a) the explicit differentiation between intraindividual variation, interindividual differences, and interindividual differences in intraindividual change; (b) stochastic models of developmental change; (c) mathematical representation of growth curves; (d) structural equation models; (e) analysis of variance applications, both univariate and multivariate; and (f) analysis of categorical data.

PREFACE

This report to the National Institute of Education represents the core of an advanced manual being developed to aid the design and conduct of longitudinal research in the behavioral and social sciences in general with particular emphasis on developmental/educational concerns. The various chapters presented here, as well as a few additional ones to be included in a projected commercially available volume, cover aspects of formulating research questions that are inherently longitudinal, designing appropriate empirical research studies, and analyzing effectively the resulting data. Our primary concern in choosing topics and selecting authors has been to provide this manual as a common meeting place, as it were, between leading experts in particularly salient aspects of longitudinal research and empirical researchers whose capacity for asking keen and interesting substantive questions exceeds their familiarity with longitudinal research tools presently at the "cutting edge" of methodological innovation.

Presented in Appendix A is a prospectus representing the contents of the entire manual as it is currently envisioned. Negotiations to secure a commercial publisher are currently underway and will be completed soon. The investigators have also completed a major part of the introductory and overview material to be included in the commercial volume. Appendix B includes a list of project related papers by the investigators that were developed during the course of the contract period and credited accordingly.

We wish to acknowledge the work and ideas of several people who helped to make this project possible. In addition to various erstwhile

and present personnel of the National Institute of Education whose helpful advice was greatly appreciated, we are appreciative of the advice and counsel of our colleagues here in the College of Human Development at The Pennsylvania State University concerning potential contributors and pertinent references. Perhaps most importantly, our collaboration with the contributors to this report was not only enjoyable and productive but also educational for us.

A great deal of gratitude is also due our project staff. Our graduate assistants, especially Steven W. Cornelius and Allison Okada Wollitzer, not only offered valuable intellectual input but were extremely helpful in tracking down references and assisting with editing. T. J. Winand and Richard L. Erwin of the Institute for the Study of Human Development helped us repeatedly in the management of resources. An expression of profound appreciation is due several very capable project and divisional secretaries and typists who assisted at various stages on the project, including: Sally Barber, Diane Bernd, Jo Ann Christina, Kathy F. Droskinis, Kathy Hooven, Miriam Landsman, Joy Lose, Patty Senior, and Ingrid Tarantelli. To all of these and others whom we may have neglected to mention -- thanks!

May, 1976
John R. Nesselroade
Paul B. Baltes

Towards a Unified Framework for Psychometric Concepts in the Multivariate Developmental Situation: Intraindividual Change and Inter- and Intraindividual Differences

by

Allan R. Buss

University of Alberta

Table of Contents

TOWARDS A UNIFIED FRAMEWORK FOR PSYCHOMETRIC CONCEPTS IN THE MULTIVARIATE DEVELOPMENTAL SITUATION: INTRAINDIVIDUAL CHANGE AND INTER- AND INTRAINDIVIDUAL DIFFERENCES

Allan R. Buss

University of Alberta

ABSTRACT

I. INTRODUCTION

II. OVERVIEW OF THE GENERAL DEVELOPMENTAL MODEL

III. PSYCHOMETRIC CONCEPTS AND THE GENERAL DEVELOPMENTAL MODEL: STABILITY

IV. PSYCHOMETRIC CONCEPTS AND THE GENERAL DEVELOPMENT

V. CONCLUSION

VI. SUMMARY

Abstract

A general developmental model for considering interindividual differences, intraindividual differences, and intraindividual change is presented. The model consists of defining each of the latter concepts in terms of Person X Variable X Occasion data and considering their interrelationships. Two major psychometric concepts, stability and regression, are then singled out for special consideration within the model. The latter tack provides for some useful conceptual distinctions, where several different kinds or aspects of both stability and regression become readily apparent. In this way, the developmental researcher can become more self-conscious of, and achieve greater clarity of, some key psychometric issues which undoubtedly will intrude upon his theory and practice.

I. INTRODUCTION

The mission of this brief chapter is to present the beginnings of a unified conceptual framework from which to view traditional psychometric concepts such as stability, regression, reliability, and measurement error within a developmental context. In addition, the emphasis will be upon the multivariate situation--accommodating the notion of repeated measures through time on multiple variables for several individuals. Critical in the model to be presented are the concepts of interindividual differences, intraindividual differences, intraindividual changes, and their interrelationships. To the extent that developmental psychologists should focus on intraindividual change and interindividual differences in intraindividual change (Wohlwill, 1970, 1973; Baltes, 1973), then there is a real need to consider within developmental methodology the interface between such concepts as intraindividual change and inter- and intraindividual differences, and traditional psychometric issues. Thus what follows is an exercise in concept methodology rather than in formal methodology. In this way, the practicing developmental psychologist may acquire a firmer grasp of some of the psychometric concepts necessary for carrying out valid developmental research in the multivariate situation.

II. OVERVIEW OF THE GENERAL DEVELOPMENTAL MODEL

In 1974 I proposed a general developmental model (Buss, 1974a) which attempted to integrate the concepts of interindividual differences, intraindividual differences, and intraindividual change within a multivariate developmental perspective. The major focus at that time was upon data-gathering strategies and data analyses rather than on traditional psychometric issues. In order to prepare the ground for a consideration of the latter, a brief overview of the 1974 model will be undertaken at this time. Note at the outset that the general developmental model proposed by Buss (1974a) is not identical with the one advanced by Schaie (1965) in the context of age-cohort research.

A. Multiple Values on One Dimension

The basic general developmental model a la 1974 involved taking Cattell's (1946, 1952) three-dimensional Person X Variable X Occasion covariation chart as a basis for defining the concepts of interindividual differences, intraindividual differences, and intraindividual change and there interrelationships. Figure 1 presents the simple case, where the datum in each cell is an individual's score on a variable at a particular occasion. In this model, interindividual differences are defined by sampling across individuals for each variable at one occasion; intraindividual differences are defined by sampling across variables for each individual at one occasion; and finally, intraindividual changes are specified by sampling across occasions for each variable for one individual.

Insert Figure 1 about here

B. Multiple Values on Two Dimensions

Figure 2 extends the concepts of interindividual differences, intraindividual differences, and intraindividual changes by considering the six possible ways of comparative sampling across each of the three dimensions. That is to say, for each of the three dimensions, the simple case is indicated in which at least two values, components, or "ids" (Cattell, 1966) are sampled across each of the remaining two dimensions or sets. The six cases thus generated are: (a) interindividual differences in intraindividual differences, in which individuals are compared in terms of sampling across variables at one occasion; (b) intervariable differences in interindividual differences, in which variables are compared in terms of sampling across individuals at one occasion; (c) interoccasion differences (changes) in intraindividual differences,

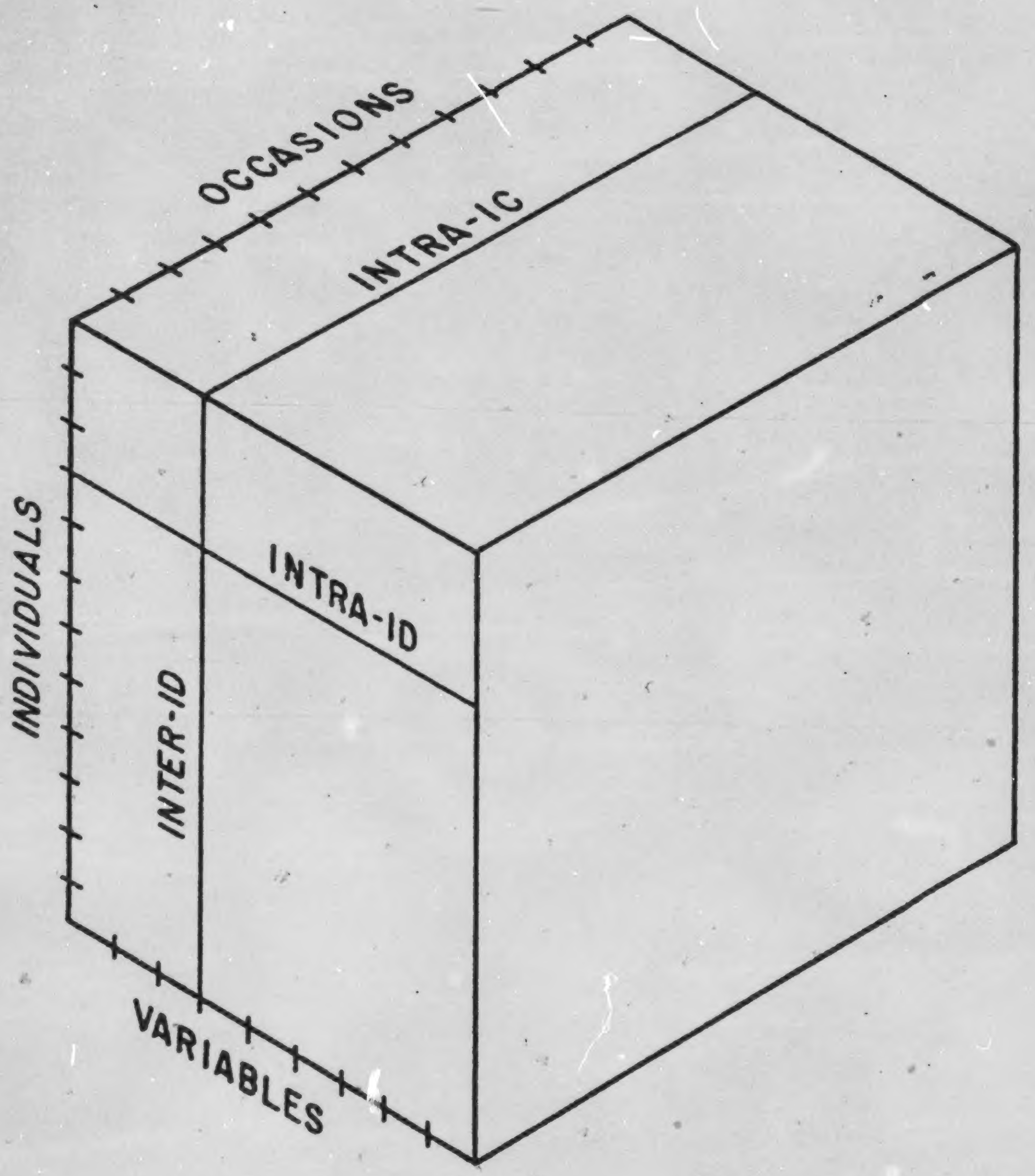

OCCASIONS
INTRA-IC
INDIVIDUALS
INTER-ID
INTRA-ID
VARIABLES

Figure 1 -- The three cases generated by sampling across each of the three dimensions of individuals, variables, and occasions are interindividual differences (Inter-ID), intraindividual differences (Intra-ID), and intraindividual changes (Intra-IC), respectively. (from Buss, 1974)

in which occasions are compared in terms of sampling across variables for one individual; (d) intervariable differences (or intraindividual differences) in intraindividual changes, in which variables are compared in terms of sampling across occasions for one individual; (e) interindividual differences in intraindividual changes, in which individuals are compared in terms of sampling across occasions for one variable; and finally, (f) interoccasion differences (changes) in interindividual differences, in which occasions are compared in terms of sampling across individuals for one variable.

Insert Figure 2 about here

Each of the above six data-gathering strategies is defined by what is compared--which gives the first aspect or the interindividual differences, intervariable differences, or interoccasion differences part--and in terms of what set is sampled across--which gives the second part of interindividual differences, intraindividual differences, or intraindividual changes aspect. The two interoccasion comparison cases (see c and f above) may be considered as changes through time. It may be appropriate to consider the (d) case as intraindividual differences in intraindividual changes, since different variables are compared in terms of intraindividual changes. This observation reveals that there are two distinct ways of operationalizing the concept of intraindividual differences in Figure 2: by sampling across variables for one individual at one occasion (a and c) and by comparing variables in terms of sampling across occasions for one individual (d). Similarly, there are two distinct views of interindividual differences: by sampling across individuals for each variable at one occasion (b and f), and by comparing individuals in terms of either sampling across variables at one occasion (a), or by sampling across occasions for one variable (e).

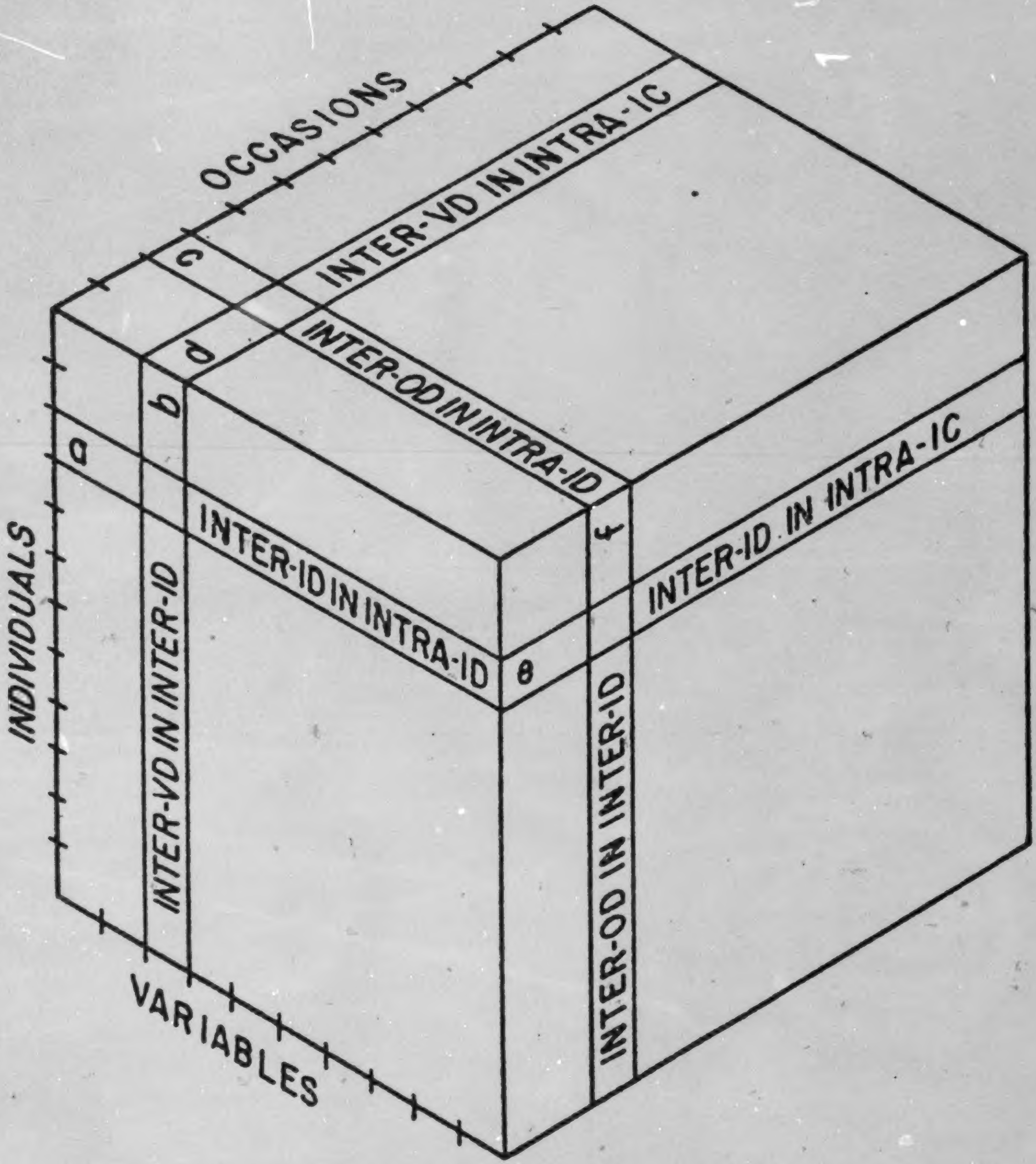

OCCASIONS
INDIVIDUALS
VARIABLES
a
b
c
d
e
f
INTER-VD IN INTRA-IC
INTER-OD IN INTRA-ID
INTER-ID IN INTRA-IC
INTER-ID IN INTRA-ID
INTER-VD IN INTER-ID
INTER-OD IN INTER-ID

Figure 2 -- The six cases generated by comparative cross-sampling for the simple case in which only two components from one dimension are compared in terms of sampling across a second dimension are shown. (Abbreviations for terms are as follows: ID=individual differences, IC=individual changes, VD=variable differences, and OD=occasion differences.) (from Buss, 1974)

Each of the six cases in Figure 2 is illustrated by comparative sampling for two ids, although the more general case would involve an entire two-dimensional matrix. The latter situation would make use of the exact same data from an Individual X Variable matrix at one occasion, but in (a) one would first get an overall measure (e.g., a variance measure) of variable differences within each individual (intraindividual differences) and compare individuals (interindividual differences in intraindividual differences) in terms of variances, while in (b) one would first get a variance measure of differences between individuals for each variable (interindividual differences) and compare variable variances (intervariable differences in interindividual differences). *Mutatis mutandis*, a similar situation exists for the remaining four cases in Figure 2.

Since the (a) and (b) data-gathering strategies are carried out at one occasion, they are not particularly useful for addressing developmental aspects of changes in variable scores, unless one simultaneously considers the third dimension of occasions (see below). In contrast, the remaining four data-gathering strategies, which involve, in part, sequential dependent measures or changes in variable scores through time, would be especially useful developmental paradigms. In considering the two cases (d) and (e), which are samples across the occasion dimension and are thus concerned with intraindividual changes, one must focus on comparing either variables (d) or individuals (e), and this requires that the unit for analysis be the plotting of the entire set of variable scores through time. A useful statistical technique that could be employed here would be testing for trends. This focus on the pattern of changes in variable scores may be contrasted with the other four cases, in which the unit of analysis for making comparisons would be variances. In the (a) case, for example, where interindividual differences or comparisons are made with respect to intraindividual differences, the appropriate index for the latter is a measure of the within-person variance of variable scores at one occasion. In order to compare variables at one occasion in terms of interindividual differences (b), again it is a variance measure that captures the extent of the interindividual differences. A similar situation holds for cases (c) and (f).

Variable scores are typically standardized across individuals for each variable at one occasion. This common practice needs to be avoided in the present scheme, since it would result in identical variances for each variable at each occasion (the variance of a standardized variable is equal to unity). If such a standardizing procedure were adopted, it would be impossible to detect intervariable differences in interindividual differences (b). By a similar argument, it would be undesirable to standardize each variable across occasions for each person, each occasion across variables for each individual, etc. What is necessary for meaningful comparisons for all six cases is to standardize each variable in terms of both individuals and occasions, that is, across each rectangular "slab" or matrix for each variable. In this way, spurious, identical interindividual differences variances for each variable at each occasion, which are brought about by rescaling procedures, are avoided. One of the advantages of standardizing in the manner being recommended here is that absolute changes in variable scores result in correspondingly higher or lower standard scores, since variables are not restandardized within each occasion.

C. Multiple Values on Three Dimensions

It is possible to extend each of the six data-gathering strategies outlined above in that situation in which one also samples through the third dimension. In other words, there is a three-step process here, in which one first samples across the first dimension, compares such sampling in terms of the ids or components on the second dimensions (the six cases just outlined above), and then proceeds to sample the comparisons of cross-sampling through the third dimension. In the complete three-step procedure, two dimensional "slabs" or matrices are compared. This process generates six cases, in which those data-gathering strategies in Figure 2 are now moderated by a term referring to the third dimension that is sampled through.

Insert Figure 3 about here

Figure 3 illustrates the complete three-step procedure for the simple case in which two ids (which have been sampled across on one dimension) are successively compared across the third dimension. The more general case would involve successive sampling of entire matrices through the third dimension. As before, the latter procedure could involve the same data for various cases, but it is the operational sequence of the three sampling steps that determines the six separate relational systems.

The six, three-step cases in Figure 3 are the following: (a) interoccasion differences (changes) in interindividual differences in intraindividual differences, in which the variances reflecting the extent of intraindividual differences for each individual at an occasion are compared for individuals through time or the occasion dimension; (b) interoccasion differences (changes) in intervariable differences in interindividual differences, in which the variances reflecting the extent of interindividual differences for each variable at an occasion are compared for variables through time or the occasion dimension; (c) interindividual differences in interoccasion differences (changes) in intraindividual differences, in which the variances reflecting the extent of

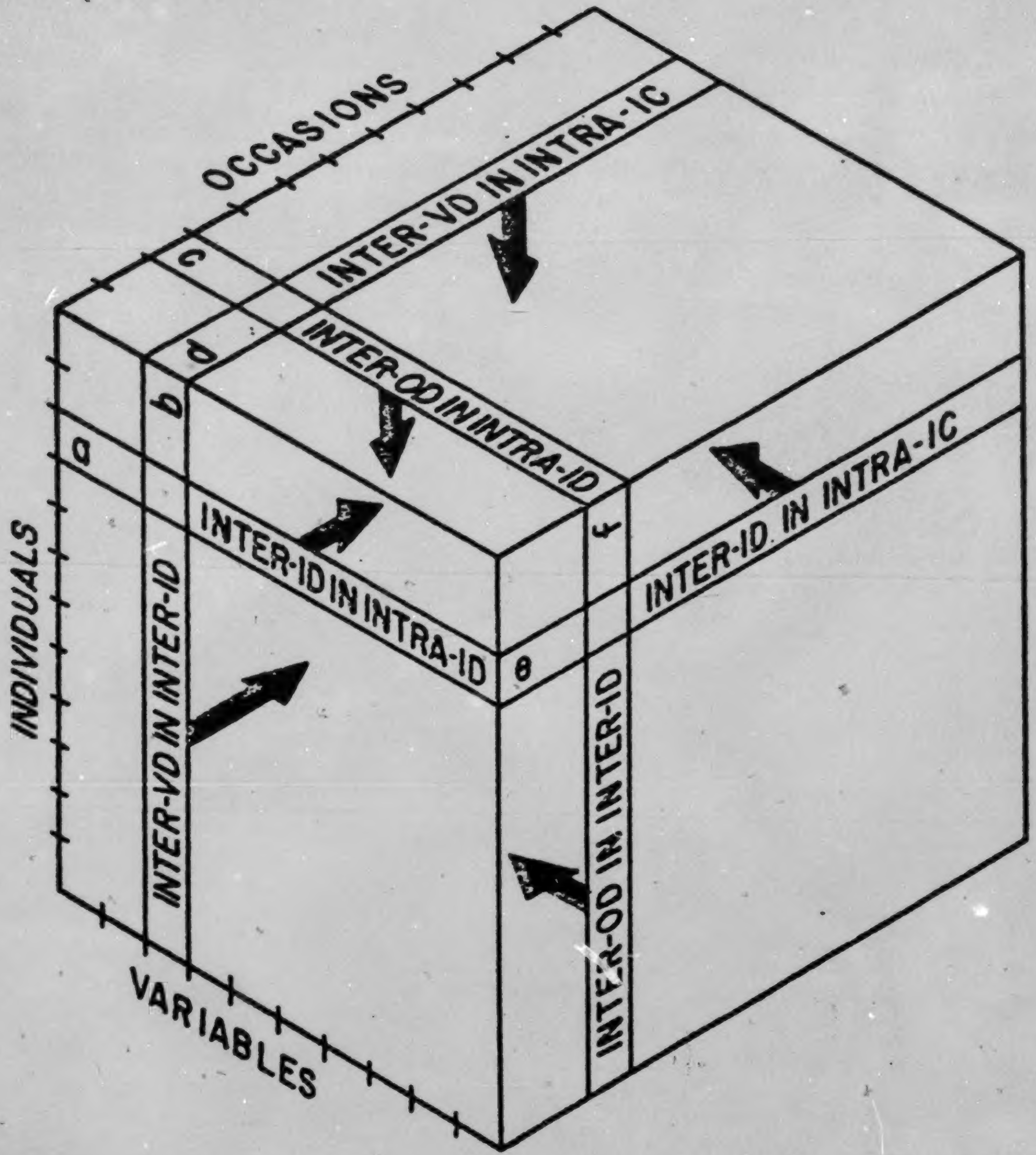

OCCASIONS
INDIVIDUALS
VARIABLES
a
b
c
d
e
f
INTER-VD IN INTRA-IC
INTER-OD IN INTRA-ID
INTER-ID IN INTRA-ID
INTER-VD IN INTER-ID
INTER-ID. IN INTRA-IC
INTER-OD IN INTER-ID

Figure 3 -- The six cases generated by sampling the labeled comparative cross-samplings through the third dimension are shown. (The simple case is indicated in which only two cross-samplings are successively compared through the third dimension. Abbreviations for terms are as follows: ID=individual differences, IC=individual changes, VD=variable differences, and OD=occasion differences.) (from Buss, 1974)

intraindividual differences for each occasion for an individual are compared for occasions through the individual dimension; (d) interindividual differences in intervariable differences (intraindividual differences) in intraindividual changes, in which the plots of variable scores across occasions at an individual are compared for variables through the individual dimension; (e) intervariable differences in interindividual differences in intraindividual changes, in which the plots of variable scores across occasions at a variable are compared for individuals through the variable dimension; and finally, (f) intervariable differences in interoccasion differences (changes) in interindividual differences, in which the variances reflecting the extent of interindividual differences for each occasion at a variable are compared for occasions through the variable dimension.

Although these extended six data-gathering strategies may appear quite complex prima facie, acquiring a firm conceptual understanding of them may be facilitated by working backward through the three steps and, consequently, their verbal designations. For example, in the case of interoccasion differences in interindividual differences in intraindividual differences, the focus is initially on the extent of intraindividual differences in variable scores at one occasion and for one individual as reflected by a variance measure. If one were then to proceed to compare such variances for two individuals at one occasion, we would arrive at the two-step concept of interindividual differences in intraindividual differences. Considering now the third dimension of occasions, in which individuals are now compared through time (occasions) in terms of the extent of intraindividual differences in variable scores, we arrive at the three-step concept of interoccasion differences in interindividual differences in intraindividual differences. One may work backwards in a similar fashion for each of the six extended cases in order to fully grasp their conceptual significance.

The 15 data-gathering strategies for interindividual differences, intraindividual differences, and intraindividual changes are summarized in Table 1

Insert Table 1 about here

----

in the order presented above, in which the first three cases consider multi-ids on only one dimension, the next six cases on two dimensions, and the last six cases on all three dimensions.

D. Limitations of the Model

Before leaving the general developmental model per se for a consideration of some key psychometric concepts within its framework, a few cautionary comments are in order. First, the model makes the assumption that the meanings of the constructs which the variables are measuring remain invariant across time. In other words, it is quantitative rather than qualitative or structural change (e.g., see Baltes & Nesselroade, 1973; Buss, 1974b; Nesselroade, 1970) which the model is capable of addressing, and it is therefore confined to slices of the lifespan where the invariance of one's constructs has been demonstrated.

A second limitation related to the above is that the present model is not focussed upon an analytic treatment of interbehavioral change. That is to say, the strength of the model lies in those situations where change is monitored within a given variable or variables rather than marking out interbehavioral changes involving cross-variable paths through time. The latter would also require techniques capable of analyzing qualitative change. A third, and once again related, limitation, serves to place the entire model in proper perspective. This point can be best appreciated by contrasting the model and its intent with related notions.

Table 1: Data-Gathering Strategies for Interindividual Differences, Intraindividual Differences, and Intraindividual Changes (From Buss, 1974)

Dimension 1: sample across	Dimension 2: compare on	Dimension 3: sample through	Type
Individuals	No	No	Inter-ID
Variables	No	No	Intra-ID
Occasions	No	No	Intra-IC
Variables	Individuals	No	Inter-ID in intra-ID
Individuals	Variables	No	Inter-VD in inter-ID
Variables	Occasions	No	Inter-OD in intra-ID
Occasions	Variables	No	Inter-VD in intra-IC
Occasions	Individuals	No	Inter-ID in intra-IC
Individuals	Occasions	No	Inter-OD in inter-ID
Variables	Individuals	Occasions	Inter-OD in inter-ID in intra-ID
Individuals	Variables	Occasions	Inter-OD in inter-VD in inter-ID
Variables	Occasions	Individuals	Inter-ID in inter-OD in intra-ID
Occasions	Variables	Individuals	Inter-ID in inter-VD in intra-IC
Occasions	Individuals	Variables	Inter-VD in inter-ID in intra-IC
Individuals	Occasions	Variables	Inter-VD in inter-OD in inter-ID

Note Abbreviations are the following: ID = individual differences, IC = individual changes, VD = variable differences and OD = occasion differences.

Mention has already been made of Cattell's (1946, 1952) original three-dimensional covariation chart. The purpose of that model was to set out the six different factor analytic techniques defined by what was correlated (persons, variables, or occasions) over what (one of the two remaining options after the initial selection). In the expanded version of the covariation chart into a generic data box Cattell (1966) has gone beyond his initial desire to set out various methods of factor analyzing Person X Variable X Occasion data and has employed a full 10 dimensions for indexing a datum. Important to note in the present context is that Cattell's more recent treatment of data relations is not confined to specific data analysis techniques. Implicit in his data box is the notion that different data analyses are appropriate for different aspects or relational systems. In the context of longitudinal data analysis techniques, Kowalski and Guire (1974) have explicitly linked specific analytic techniques[1] to various relational aspects of Person X Variable X Occasion data. In contrast to Cattell's (1966) comprehensive treatment of how one can conceptualize data in general, and in contrast to Kowalski and Guire's (1974) general account of the various analytic techniques available for longitudinal data, the present model is restricted to conceptualizing specific concepts (inter- and intraindividual differences, intraindividual change) within the three-dimensional data frame. Its intent, therefore, is to focus in upon a specific set of concepts as these are related to longitudinal data. The present model makes no attempt to explore terrain previously charted out by others.

III. PSYCHOMETRIC CONCEPTS AND THE GENERAL DEVELOPMENTAL MODEL: STABILITY

The two psychometric concepts singled out for specific treatment in terms of the multivariate developmental situation and the preceeding model are stability and regression to the mean. Other concepts, such as reliability, measurement error, true scores, etc., are introduced as needed and as they are related specifically to stability and regression.

A. Stability in Two Dimensions

By stability in the multivariate developmental situation, one can mean either the stability of differences between or within persons through time. In considering Figure 2, these two cases translate respectively into noting either the degree of stability of interindividual differences scores on a single variable through time (case f), or the stability of intraindividual differences scores on a single person through time (case c).

In each of these simple cases, multiple values on one dimension (either individuals or variables) for one id (either a variable or individual) are obtained on two successive occasions. The simple two-occasion situation would involve calculating a correlation coefficient as the stability coefficient, whereas the multiple occasion situation would involve calculating successive stability coefficients (see below). Since the Pearson product moment correlation restandardizes scores at each occasion, that is, it cancels out means and variances, this statistic would permit inferences concerning the degree of stability of the pattern or shape of the multiple scores through time. Thus, it would be possible to have absolute changes in scores, yet obtain a high stability coefficient so long as there was relatively high invariance of the pattern of interindividual differences through time. It can be noted in passing that the ceiling of a two-occasion product moment stability coefficient would be the lower reliability coefficient as determined by assessing reliability at each occasion.

An alternative index of stability that takes more than just the consistency of shape into account would involve one of the measures of similarity based upon distance (see Bolz, 1972, for a review of such indices of similarity). Cattell's (Cattell et al., 1966) pattern similarity coefficient is probably the superior statistic in that it simultaneously considers differences between two patterns of variable scores due to: elevation (the mean of all scores in a profile), scatter (the square root of the sum of squares of the deviation scores about the mean), and shape (the residual information after equating two patterns for both elevation and scatter). In addition, the pattern similarity coefficient is readily meaningful--varying between 0 (no pattern similarity) and +1 (perfect pattern similarity), and it has known distribution and a test for significance (Horn, 1961).

If one were to employ the pattern similarity coefficient as a two-occasion index of stability of either interindividual differences or intraindividual differences, it would be necessary to standardize across each rectangular "slab" or matrix for each variable as previously outlined. In this way, one can make the best use of the additional measurement properties of the pattern similarity coefficient, that is, its sensitivity to differences in both the means and variances of two patterns. Thus, the pattern similarity coefficient should be used as an index of the stability of inter-or intraindividual differences when the concern is for absolute rather than relative pattern invariance over two occasions.

B. Stability in Three Dimensions

The above notions of stability can be readily generalized to the three dimensional situation, that is, where it is desirable to make variable comparisons in the stability of interindividual differences, or, person comparisons in the stability of intraindividual differences. Each of these cases in the two-occasion situation can be visualized by reference to Figure 3. Variable comparisons in the stability of interindividual differences involves case (f)--where it is extended through the third dimension of Variables such that variable differences in stability coefficients are observed. The same rationale follows in making person comparisons in the stability of intraindividual differences, which involves extending case (c) in Figure 3 through the third dimension of Individuals. In either of these two cases, the product moment or pattern similarity coefficient may be used, depending upon the question being asked in light of the unique properties of each of these statistics as previously discussed.

Thus far the concept of stability has been extended to both inter-and intraindividual differences in the multiple variable and person situation. However, a truly adequate treatment of stability in a multivariate developmental context must provide for more than the restrictive two-occasion model. Extending cases (f) and (c) in Figure 2 right across occasions to involve an entire matrix or "slab," illustrates successive or multiple occasion data-gathering. Thus in case (f), successive stability coefficients for a single variable could be obtained, where each stability coefficient is calculated from adjacent occasions. The values of the stability coefficients could then be plotted for purposes of variable comparisons in trends over time, and Figure 4 illustrates three prototype cases: increasing stability (IS), stable stability (SS), and decreasing stability (DS). By using curve fitting techniques, one may then determine variable differences in stability trends through time. The exact same logic may be applied to case (c) in Figure 2, where Figure 4 may now be regarded as illustrating three prototype stability functions for intraindividual differences. Each function, then, characterizes a particular individual rather than a particular variable. The term "prototype" in the latter situation is more than a mere casual use of labels, since it may indeed be possible to distinguish "types" or clusters of persons on the basis of stability functions for intraindividual differences.

Insert Figure 4 about here

What about the interrelationship between the degree of stability of interindividual differences and the degree of stability of intraindividual differences, where the same data are analyzed in the appropriate manner? High stability of interindividual differences on each of the variables is a prerequisite for high stability of intraindividual differences for each person, and vice versa. In other words, in the extreme case (i.e., considering the stability of the entire Variable X Individual matrix through time), both the stability of inter- and intraindividual differences are two different ways of looking at the same phenomenon. However, the direct translation between the degree of stability of inter- and intraindividual differences evaporates as soon as one selectively focuses upon a subsample of variables or persons through a limited number of occasions. The latter statement, of course, does not deny the ever present nonindependence of the stability of inter- and intraindividual differences when calculated from the same data.

C. Stability of Population Parameters

Having distinguished between stability of inter- and intraindividual differences, there remains another important sense of the term "stability" applicable to the multivariate developmental situation. Thus, one may speak of the stability of various variable properties of a population through time,

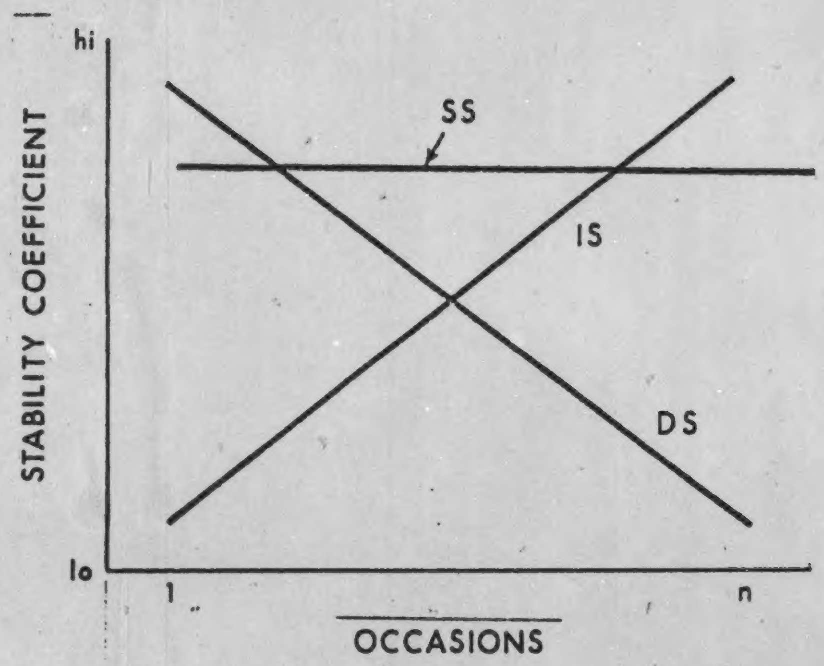


Figure 4 -- Three prototype cases of increasing stability (IS), stable stability (SS), and decreasing stability (DS).

where the focus is now upon the degree of stability of certain descriptive statistics such as the mean or variance. Clearly, in this case, one would not restandardize the variable scores at each occasion, since this would mask changes in absolute value of the mean and variance. In mapping out various statistical properties of variables over time, it is quite possible to have stability of, say, the mean and variance, yet, at the same time, have radical instability of inter- and/or intraindividual differences.

The distinction being made here is conceptually quite important, since the stability or "dynamic equilibrium" of a population on a particular variable does not entail stability and/or systematic change at the individual level. To the extent that in multivariate developmental psychology there should be an emphasis upon intraindividual change and interindividual differences in intraindividual change (e.g., in the present context, the degree of stability of intraindividual differences and person differences or interindividual differences in the stability of intraindividual differences), then it is important to make the kind of distinction presently being made. This is not to say that questions concerning the stability of properties of a population on a particular variable, or the stability of interindividual differences through time, may not be extremely interesting or important. Rather, and this is the basic point, there are several ways one can frame questions concerning stability in the multivariate developmental situation, and the practicing researcher should be quite conscious of the various alternatives available and effect an adequate match between the research question and the concept of stability employed.

IV. PSYCHOMETRIC CONCEPTS AND THE GENERAL DEVELOPMENTAL MODEL: REGRESSION

A. Regression as a Prediction Model versus Type of Change

Regression towards the mean is a thorny issue in developmental psychology, and, although it has been discussed within a developmental context (e.g., Baltes & Nesselroade, 1976; Baltes, Nesselroade, Schaie, & Labouvie, 1972; Clarke, Clarke, and Brown, 1960; Furby, 1973), there remains a need for further clarification. Any discussion of regression must, of necessity, consider such psychometric concepts as measurement error, reliability, true scores, and so on. In discussing regression within a developmental context, it is advantageous, as suggested by Baltes & Nesselroade (1976), to separate the typical psychometric issue associated with this term, namely, regression as a prediction model, from observed regression in the data.

The previously outlined developmental model of Buss (1974a) can be fruitfully consulted in attempting to keep separate regression as a prediction model and observed regression in the data. Thus in Figure 2, regression as a prediction model can be best located in terms of case (b), where one would employ regression techniques for predicting scores on one variable from scores on another variable. Regression in this instance is not "in" the data, or in other words, is not a "real" phenomenon.

In order to better understand the point I am trying to make here, consider case (f) in Figure 2, where measures are obtained on the same variable on two separate occasions. In the latter instance, there is continuity over time, and if there is regression from occasion to occasion, it is a phenomenon intrinsic to these data. In other words, and this is the basic point, when regressing one variable against another different variable, we are using regression as a prediction model. When we obtain repeated measures for two occasions on the same variable, regression toward the mean, if it occurs, is a phenomenon to be explained. In the latter situation, regression toward the mean is a special kind of change in one's data. Should one desire, variable differences in regression as a type of change can be located in the general model by considering case (f) through the third dimension, that is, as illustrated in Figure 3.

In the following discussion, regression as a prediction model is bracketed. The focus of attention is exclusively upon that situation where observed regression is in the data (regression as a type of change) in order to try to clarify some of the developmental issues in this area.

B. Regression of Extreme Samples

Observed regression to the mean may be associated with either changes in true scores or changes in error scores in the classical measurement model of $\underline{x} = \underline{t} + \underline{e}$, that is, the observed score is the sum of a true score component and an error score component. Important to note, then, is that observed regression, which is in the data, may or may not reflect true changes in the underlying trait or variable of a particular sample.

The developmental psychologist must, in certain situations to be explicated, take account of observed regression associated with the error part of scores in order to make valid inferences concerning real or true score change over time. The classical case where the developmental psychologist simply must consider the extent to which observed regession towards the mean is associated with true changes in the variable is illustrated in Figure 5 (see also Baltes & Nesselroade, 1976; Furby, 1973).

In Figure 5, the sample of individuals is initially selected from the extreme range of the normal distribution of the population. We can make the

----

Insert Figure 5 about here

----

assumption in Figure 5 that measurement error exists and is constant across occasions, that is, reliability is invariant. As the second occasion, the

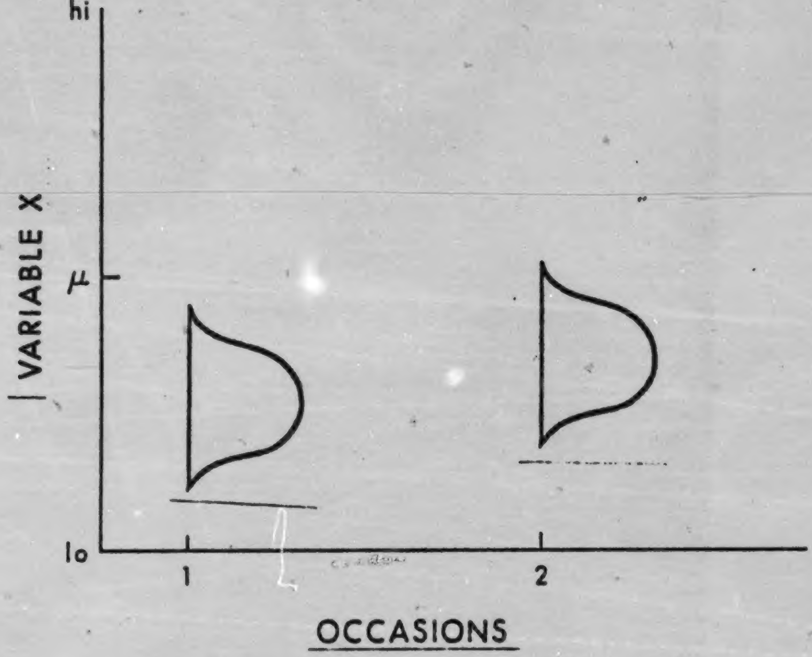

hi
VARIABLE X
μ
lo
1
2
OCCASIONS

Figure 5 -- The classical situation of regression towards the population mean, where the sample is initially selected from the extreme range of the normal distribution of the population.

entire sample has shifted closer to the population mean $\underline{u}$. To the entent that the reliability of variable $\underline{x}$ is less than unity, that is, there is some measurement error, then there will be observed regression associated with changes in error scores. This follows from the classical measurement model--a model that forms the bedrock of measurement theory--and a model which must be accommodated until displaced by something shown to be superior.

The classical measurement model states that (because errors of measurement are assumed to be uncorrelated over time) extreme scores in a distribution are biased in that same extreme direction, and that on a second occasion, the error associated with these individuals' scores will, on the average, be less--thereby effecting observed regression towards the population mean. Another way of saying this is that in nonerror-free measures, observed scores are biased estimates of true scores, where observed scores above the population mean are biased upwards, and vice versa for observed scores below the population mean.

For the developmental psychologist who is working with a select sample of nonerror-free extreme scores, regression toward the mean associated with changes in error scores will occur. It will show up as observed regression, all other things being equal. To the extent that there are true changes in the underlying trait or construct which a variable is tapping, this may effect no observed change in the sample distribution if the true score and error score changes, on the average, cancel each other out. Thus, the developmental researcher, when working with a sample under conditions described above, must know from where the sample came, as well as the reliability of measures, in order to make valid conclusions as to the extent of true score or "real" changes in the underlying construct over time. The latter conclusion is not the consequence of a specific prediction model, but is the consequence of a very general measurement model underlying all measurement--psychological or otherwise.

In regard to integrating the concept of regression within the present methodological model, we can note that, in that ideal case of error-free measures and where regression is associated with changes in true scores and, for arguments sake, the Pearson product moment stability coefficient is unity, then in the shift towards the population mean in Figure 5 there will be no interindividual differences in intraindividual change. The latter would be a very unlikely situation in any actual research outcome. More probable would be observed interindividual differences in intraindividual change towards the population mean associated with either real or/and spurious factors, respectively paralleling true score and error score changes.

C. Regression of Representative Samples

There is another, quite different, situation involving regression toward the mean, and previous discussions have failed to place it in sharp contrast with what is most typically meant by regression (but see Baltes & Nesselroade, 1976). On occasion there has even been a tendency to confuse the two situations of regression, or at least switch from one to the other without adequate awareness.

The second major case of regression toward the mean is illustrated in Figure 6 where, in this instance, the sample distribution can be considered as representative of a population rather than coming from a selective region of the population distribution as was previously the case. In Figure 6 there is observed regression towards the mean, but in this instance, regression is defined by a reduction in the sample variance rather than a shift in the sample mean (and thus the sample) towards the population mean. Thus, in Figure 6 the means at each occasion are identical yet there is within-sample regression towards the mean, and within-population regression to the mean to the extent the sample is representative of a population. As before, such observed regression may be associated with changes in either true scores or changes in error scores. However, in this second case of regression, any regression associated with changes in error scores is due to increased reliability across occasions and a reduction of the error variance--which was not the case in the first instance of regression where it was assumed reliability was constant over time. Of course, in considering only changes in error scores, both within-sample regression to the sample mean and regression of the sample to the population mean could occur simultaneously in Figure 5 if there were an increase in reliability over time.

Insert Figure 6 about here

In Figure 7 it can be seen that mean sample changes in either an upwards or downwards direction are independent of within-sample regression to the mean. Thus the entire distribution of scores may shift in either direction--regression still being defined as a reduction in variance. The latter observation brings us to the point that, in the multiple occasion situation, changes in regression to the mean may be mapped via noting changes in variance through time. However, this conclusion reveals that observed regression is "merely" observed change in a specified direction, although there is more to it than that. The observed change may be associated with changes in either error scores (reliability) or/and true scores, and valid developmental conclusions require separating these two sources in both types of regression discussed in this section.

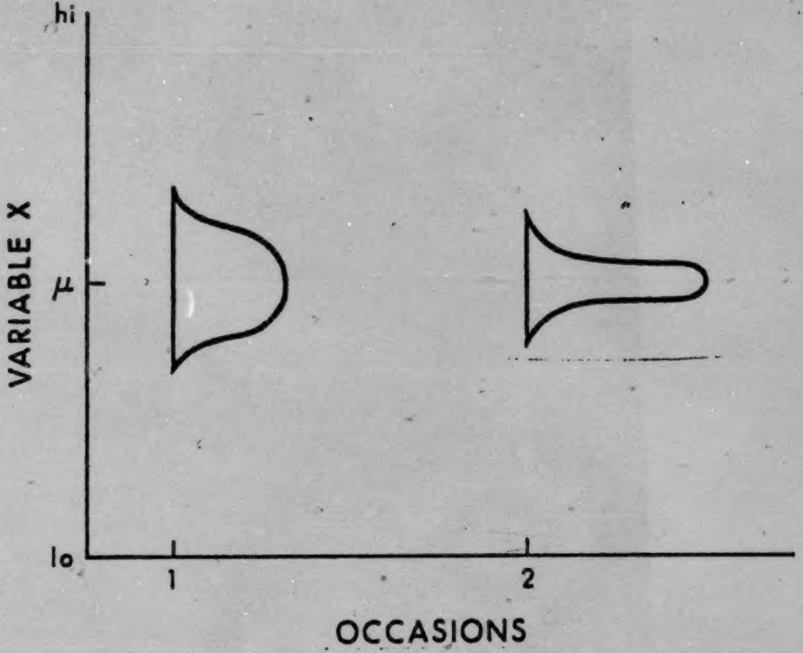

hi
VARIABLE X
μ
lo
1
2
OCCASIONS

Figure 6 -- Regression towards the population and sample mean where the sample is representative of the population.

Insert Figure 7 about here

D. Concluding comments on Regression

While is is true that within-sample regression may or may not occur--depending upon whether there is a reduction in error variance and/or real changes in true scores towards the mean--regression of observed scores, as discussed in the first instance and illustrated in Figure 5, will occur, assuming nonerror-free measures and all other things being equal. Developmental psychologists will need to assess any true score change in the sample mean either towards or away from the population mean in light of error score change.

Recently, Baltes and Nesselroade (1976) have concluded that regression toward the mean is often an irrelevant issue in developmental research to the extent that we should focus upon change and multiple occasion data beyond the two-occasion case. According to them, regression is only one form of change rather than some immutable law and in multiple occasion data, it can be assessed via error-centered baseline comparisons. This view has much to recommend it, since it places the typical two-occasion regression situation within a broader framework for the multivariate developmental researcher. Such being the case, the spirit of their effort is consistent with perspectives developed here.

V. SUMMARY

The concepts of interindividual differences, intraindividual differences, and intraindividual change were defined in terms of sampling across one of the three dimensions of individuals, variables, and occasions, respectively. Each of these concepts was then considered in comparative sampling by introducing a second dimension, thus generating six data-gathering strategies. Each of

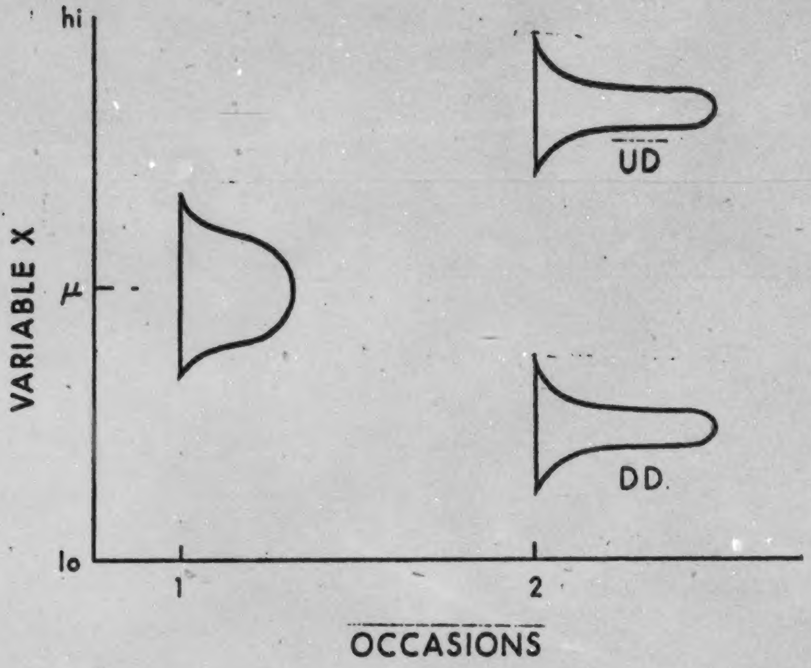

hi
lo
μ
VARIABLE X
UD
DD
1
2
OCCASIONS

Figure 7 -- Regression towards the sample mean but not towards the population mean in two different situations: change in an upwards direction (UD) and change in a downwards direction (DD).

the latter six cases was extended to include the third dimension through which the comparative sampling is sampled. Out of the total 15 data-gathering strategies considered, 11 were defined in part by the occasion dimension and are therefore capable of dealing with change data.

Limitations of the general developmental model were discussed. These included: (a) the model rests upon the assumption that the meaning of the constructs which the variables are measuring remain invariant; (b) the model is not equipped to deal in an analytic way with interbehavioral change; and finally, (c) the model is to be used for focussing upon a specific set of concepts as these are related to longitudinal data, rather than for attempting a general and comprehensive deliniation of various developmental data analysis techniques.

Having outlined the general developmental model and its limitations, two important psychometric concepts were considered within the developed framework: stability and regression. Two kinds of stability were outlined which were defined by two of the three dimensions of persons, variables, and occasions: (a) the stability of interindividual differences on a single variable through time; and (b) the stability of intraindividual differences on a single person through time. Stability in three dimensions involves extending each of these types of comparing stability coefficients across variables or persons respectively.

Stability can also involve multiple occasion data, where successive stability coefficients are calculated at each occasion for determining changes in stability of either interindividual differences on one variable, or intraindividual differences for one person. One could then make variable or person comparisons, respectively, of trends in stability functions. Finally, stability of population parameters was another aspect considered, where it was pointed out that one could plot the values of certain descriptive statistics over occasions. This meaning of stability reveals the important conceptual point that, it is quite possible to have stability of, say, the mean and variance of a population, yet, at the same time, have radical instability of inter- and/or intraindividual differences.

In considering regression toward the mean, two fundamental types were defined within the general developmental model: (a) regression as a prediction model; and (b) regression as a kind of change. The former was seen as essentially nondevelopmental, and thus discussion was confined to regression as a special kind of change over time.

Two kinds of regression toward the mean as change were considered for the two occasion case only: (a) regression of extreme samples; and (b) regression of representative samples. Regression of extreme sample means to population means was seen to occur given the assumptions of the classical measurement model and all other things being equal. To the extent that there are changes in the true score component of observed scores, this will effect the amount of observed regression which would otherwise be due solely to changes in measurement error over occasions. Regression of representative samples of particular populations was seen to involve a decrease in variance over time rather than changes in sample means. Thus there can be a decrease in sample variance (and thus regression towards the mean) brought about by changes in either the true scores and/or the error scores. The mean may or may not change over time--regression in this case being independent of such change.

In conclusion it should be noted that greater clarity of psychometric concepts in the developmental situation is possible and necessary for those practicing the craft of developmental research. The proposed formulation of a general developmental model involving interindividual differences, intraindividual differences, and intraindividual change is helpful in this regard.

Footnotes

[1] Some of those considered are the following: curve fitting, factor analysis, multivariate analysis of variance, polynomial growth curve models, regression, time series, etc.

Mathematical Representations of Development Theories

by

Burton Singer
Columbia University

Seymour Spilerman
University of Wisconsin-Madison

Table of Contents

MATHEMATICAL REPRESENTATIONS OF DEVELOPMENT THEORIES

Burton Singer — Columbia University

Seymour Spilerman — University of Wisconsin-Madison

ABSTRACT

I. INTRODUCTION

II. THE CONCEPT OF DEVELOPMENT STAGES AND A MATHEMATICAL FORMULATION OF STAGE PROGRESSIONS

III. MODELS OF MORE ELABORATE STAGE THEORIES

IV. ALTERNATIVE SPECIFICATIONS OF THE SIDE CONDITIONS

V. CONCLUSIONS AND SUMMARY

The work reported here was supported by National Science Foundation grants SOC76-17706 at Columbia University and SOC76-07698 at University of Wisconsin-Madison. Assistance was also provided by the Institute for Research on Poverty at the University of Wisconsin.

ABSTRACT

In this paper we explore the consequences of particular stage linkage structures for the evolution of a population. We first argue the importance of constructing "dynamic" models of developmental theories and show through a series of examples the implications of various stage connections for population movements. In discussing dynamic models, one thrust of our comments is to identify the sorts of process features about which assumptions must be made in order to convert a static theory about stage connections (the sort of specification commonly presented in life-span psychology) into a dynamic model. A second focus of our discussion concerns inverse problems: how to utilize a model formulation so that the stage linkage structure may be recovered from survey data of the kind collected by developmental psychologists.

MATHEMATICAL REPRESENTATIONS OF DEVELOPMENT THEORIES

I. INTRODUCTION

Although time, usually in the guise of age, is a crucial variable in developmental psychology, it is the case that formal models of developmental phenomena rarely have the character of dynamic representations, in the sense of mimicking the evolution of an empirical process through time. The analytic procedures employed most extensively by life-span psychologists are factor analysis, regression, analysis of variance, scaling, clustering, and variants of these methods (see, for instance, Nesselroade & Reese, 1973). These are powerful techniques for identifying variables that are central to the course of development in a particular substantive area (e.g., intellectual maturation, acquisition of moral values). Also, when applied to panel data, the procedures can yield insights into how the salience of key variables shifts over the life cycle, or over a portion thereof (e.g., stages in infancy, youth, adulthood).

These analytic methods do not, however, lead to dynamic formulations of developmental theories, which can be useful in testing predictions from a theory about the evolution of an empirical process, or in comparing the implications of competing explanations. By a dynamic formulation we mean a representation which incorporates into the mathematics the main assumptions about a developmental phenomenon and is specified in such a way that the relevant variables, and their postulated interrelations, are functions of time or subject's age. In this sense, like the empirical process, it too constitutes an evolving system. As a simple illustration of such a model, consider the following statements of alternative evolutionary mechanisms:

(A) The growth of a process at each instant is proportional to its potential for future growth.

(B) The growth of a process at each instant is proportional to the product of its current size and its potential for further growth.

These statements might be proposed as competing explanations of the manner by which information is diffused in a population of size N. In formulation (A), it matters not how many persons y(t) know the information of concern at instant t; only those yet to hear, numbering N-y(t), are salient to the diffusion rate. If the information were propagated by a mass media source, such as radio or television, rather than by interpersonal communication, this model might apply. Formulation (B), in comparison, is consistent with a process in which those already aware of the information "infect" the uninitiated through contact and conversation. Assuming that the informed and the uninformed mix randomly, the variable governing the evolution of the process would be y(t)[N-y(t)], which measures the rate at which individuals from the two groups come into contact.

The evolutionary mechanisms, (A) and (B), can be represented by the differential equations (I.1) and (I.2), respectively,

$$\frac{dy(t)}{d(t)} = k_1[N-y(t)], \quad y(0) = 0 \tag{I.1}$$

$$\frac{dy(t)}{d(t)} = k_2 y(t)[N-y(t)], \quad y(0) = 1 \tag{I.2}$$

where k_1 and k_2 are constants which adjust for the time unit (e.g., day, year) used in the measurements.[1] Equations (I.1) and (I.2) have for solutions (I.3) and (I.4),

$$y(t) = N(1-e^{-k_1 t}) \tag{I.3}$$

$$y(t) = \frac{Ne^{Nk_2 t}}{N-1+e^{Nk_2 t}} \tag{I.4}$$

which predict the different evolutionary paths displayed in Figure 1.

Figure 1 about here

These formulations are "dynamic" in that time appears explicitly as a variable; they are process "models" in that the predicted value of y(t) evolves according to the assumptions of a particular theory. If a researcher has data on the time course of an empirical process, he could test whether equation (I.3), (I.4), or a specification of an equivalent sort best approximates his observations. By this exercise it is often possible to select among competing explanations of the mechanism underlying a developmental process. Indeed, these very models have been applied by Coleman, Katz, and Menzel (1957) to data on drug adoptions by physicians (also see Coleman, 1964, pp. 43-45). They concluded that the drug acquisition pattern by socially integrated MD's is best represented by a logistic curve (implying mechanism [B]), while isolated MD's adopt according to the constant source model (mechanism [A]), as they are influenced principally by drug advertisements in trade journals. To our knowledge, although developmental psychologists emphasize ontogenetic processes and employ the imagery of an evolutionary system, few attempts have been made to translate their theories into formal models of the above sort.

In this paper, we describe the formulation of dynamic models where the objective is to test developmental theories against data or ascertain the consequences of particular assumptions about the structure of a process. To delimit our task, we focus on the sort of mathematics that is appropriate for studying qualitative change. As a result, the tools we introduce are pertinent to theories which postulate stage sequences, a variety of explanation with considerable precedent in developmental psychology (Piaget, 1960; Kohlberg, 1968; Ausubel & Sullivan, 1970). To the degree possible we have written this paper with a view toward substantive issues and have concentrated on the

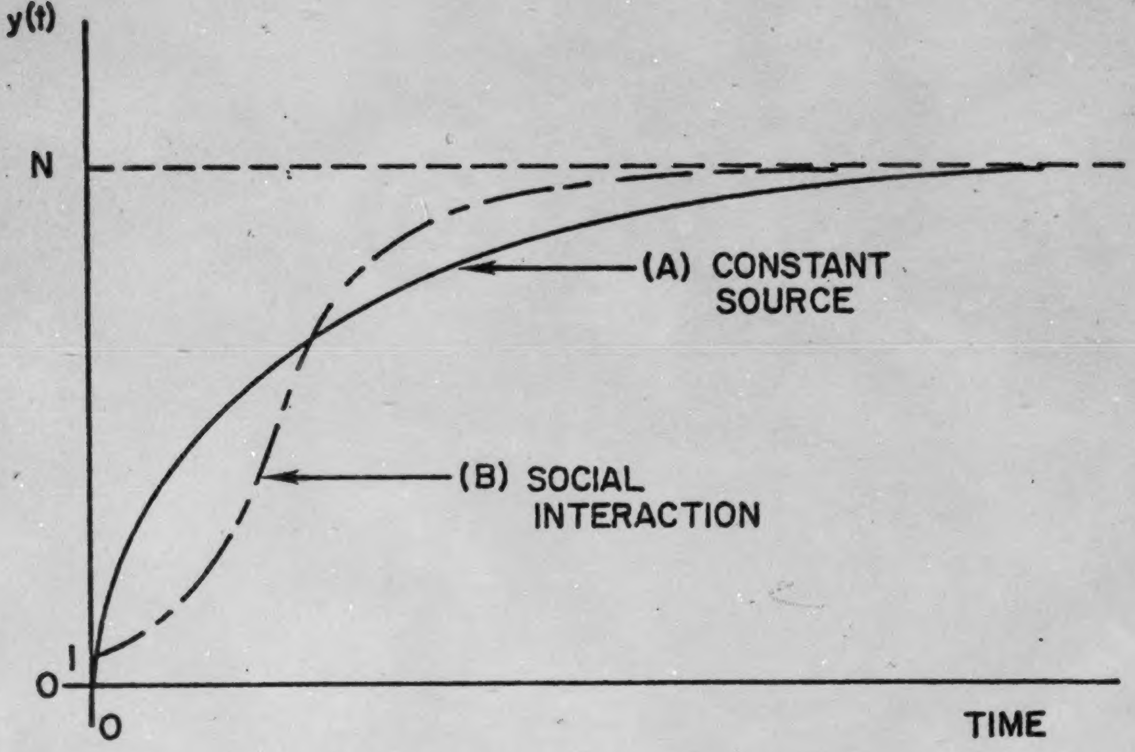


Figure 1 -- Illustrative Growth Curves for Diffusion via Social Interaction and Diffusion from a Constant Source[a].

---

[a] N = population size; y(t) = number aware of the information at time t.

translation of theoretical specifications into mathematical formalism; the reader usually is referred elsewhere for mathematical details and estimation procedures. The organization of the paper is as follows: In the next section we introduce a class of models that is suitable for studying evolutionary processes which incorporate the notion of stage. In section III we describe how particular stage theories can be cast in the framework of the general model. In section IV we relax several requirements of the basic model so that it can more realistically represent developmental phenomena.

II. THE CONCEPT OF DEVELOPMENT STAGES AND A MATHEMATICAL FORMULATION OF STAGE PROGRESSIONS

Stage sequences have been postulated for a variety of developmental processes--the evolution of moral behavior (Kohlberg, 1973), cognition (Piaget, 1954), personality (Loevinger, 1966), and motor skills (Shirley, 1933), to cite but a few topics. There also exist diverse formulations of stage models in the literature of life-span psychology. These differ with respect to the presumed sources of the stages and with regard to the rules governing movement between them. In regard to stage origins, some authors have emphasized maturational considerations, in which individuals are viewed as programmed genetically for particular behaviors or abilities to emerge (Gesell, 1954). The specification of psychosexual stages, keyed to biological activation of the sex glands, provides an illustration (Kohlberg, 1973, p. 181). Others view stages as arising from interactions with the social environment. Kohlberg (1968, pp. 1016-1024), for example, contends that experience with the cultural and physical world is necessary for cognitive stages to take the shapes they do. Still other researchers have adopted the position that stages are a useful research construct around which to discuss development, without insisting that they have an empirical existence (Kaplan, 1966; Reese, 1970).

We shall not discuss further the very important issues concerning the etiology of stages, but will focus instead on the mathematical representation of theories about stage connections and on the consequences of various linkage structures for the evolution of individuals among the stages. Formulations of stage connections in a developmental process differ according to whether the progression is viewed as unilineal or multilineal, whether stages in the sequence can be skipped, and whether regression to an earlier level is possible. A second set of considerations pertinent to the structure of developmental theories concerns the age specificity of a stage and the related matter of the variability of duration in a stage. For discussions of these topics in the context of particular substantive processes, the reader is referred to Emmerich (1968) and Kessen (1962).

To develop the mathematical apparatus for ascertaining the implications of particular stage connections, we discuss both the simplest prototype of a stage theory (for concreteness) and the general mathematical formulation.[2] Consider, then, a developmental progression consisting of n stages, in which the linkage is unilineal and there is no possibility of stage skipping or regression. An example of such a structure, with n equal to 5, is presented in panel A of Figure 2; henceforth this model is referred to as example 1. It will be convenient to also have available a matrix representation of the stage linkages. For an arbitrary n-stage structure, we define a matrix M,

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & \cdots & \cdots & m_{2n} \\ \vdots & & & \vdots \\ m_{n1} & \cdots & \cdots & m_{nn} \end{bmatrix} \qquad \text{(II.1)}$$

whose elements are m_{ij} = {probability of transferring from stage i to stage j when a transition occurs}, where $0 \le m_{ij} \le 1$, and $\sum_{j=1}^{n} m_{ij} = 1$. These restrictions on the elements of M ensure that each row of the matrix constitutes a probability distribution. We require, in addition, that $m_{ii} = 0$ for each stage i which is not an absorbing state of the process; that is, from which individuals can exit. This means we exclude the possibility of within-stage transitions, a type of move which is undefined in most developmental theories. Also, we set $m_{ii} = 1$ for each stage which is an absorbing state of the process. This is done for mathematical convenience and, as we shall see, carries no substantive implications. In the particular case of the unilineal progression (Figure 2, panel A), we have the further requirements on M: $m_{i,i+1} = 1$, and $m_{ij} = 0$ otherwise (except that $m_{55} = 1$). This matrix, M_1, is reported in panel B of Figure 2.

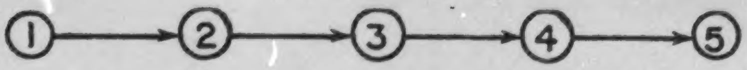
Figure 2 about here

To this point, though matrix M conveys important structural information about the process, the description of the stage progression is a static representation. To elaborate the model we must indicate how stage transition events occur. At a general level of description we assume that the time τ_k spent by an individual in stage i follows some probability distribution,

$$\text{Prob}_i(\tau_k < t \mid \tau_1, \ldots, \tau_{k-1}) \qquad \text{(II.2)}$$

where $\tau_1, \ldots, \tau_{k-1}$ report the sojourn times in earlier stages. Our imagery, therefore, is the following. An individual originates in stage i at the beginning of the process, $t_0 = 0$. He remains there for an interval τ_1, specified by a distribution function $\text{Prob}_i\ (\tau_1 < t)$, and then transfers to stage j with probability m_{ij}. He remains in this stage for a period τ_2, specified by a conditional probability distribution $\text{Prob}_j\ (\tau_2 < \tau \mid \tau_1)$, then transfers to stage k with probability m_{jk}; and so forth.[3] The process continues until some absorbing state is reached, at which point the evolution is terminated. The time path

a. Diagram of Stage Linkages



b. Matrix Representation[a] of the Stage Linkages[b]

$$M_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 2-- Representation of a Simple Unilineal Stage Structure.

---

[a] Each row of M_1 is a vector of destination probabilities. Thus, if an individual were in stage one before a transition, the row one entries would pertain and they indicate movement to stage two with probability equal to 1.

[b] The main diagonal entries are set equal to zero (with the exception of row 5) to indicate that a "move" is not defined apart from a stage transition; i.e., there is no notion of movement within a stage. The main diagonal entry of row 5 is set equal to 1 because this stage is an absorbing state ($M_{5j} = 0$ for $j \neq 5$) and the definition of m_1--see text--requires $\sum_j m_{5j} = 1$.

for the unilineal progression associated with the stage linkages of matrix M_1 is presented in Figure 3.

Figure 3 about here

Several further assumptions are necessary to complete the specification of the model. One matter concerns the relevance of an individual's past movement history to the course of his subsequent evolution among the stages. We assume,

(i) knowledge of current stage conveys all information that is relevant to forecasting future movements.

Stated technically, if $m_{ij,ab\ldots f}$ = {probability of moving from stage i to stage j at the occurrence of a transition, given prior sojourns in stages a, b,...,f}, then

$$m_{ij,ab\ldots f} = m_{ij} \, .$$

(This assumption is superfluous in the current example of a unilineal progression since there is only one possible path, but it is relevant to the evolution of a population in less restrictive models.) We indicate in the next section that this specification has been employed in descriptions of stage linkages in developmental psychology.

For an initial baseline class of models, we further assume,

(ii) the sojourn time in stage i is exponentially distributed: that is,

$$\mathrm{Prob}_i(\tau_k < t \mid \tau_1, \ldots, \tau_{k-1}) = \mathrm{Prob}_i\,(\tau_k < t)$$

$$\equiv F_i(t) = 1 - e^{-\lambda_i t} \qquad \text{(II.3)}$$

Use of the exponential distribution amounts to specifying that the probability of departing from stage i during the infinitesimal interval (t, t+dt), conditional on being in stage i at time t, equals

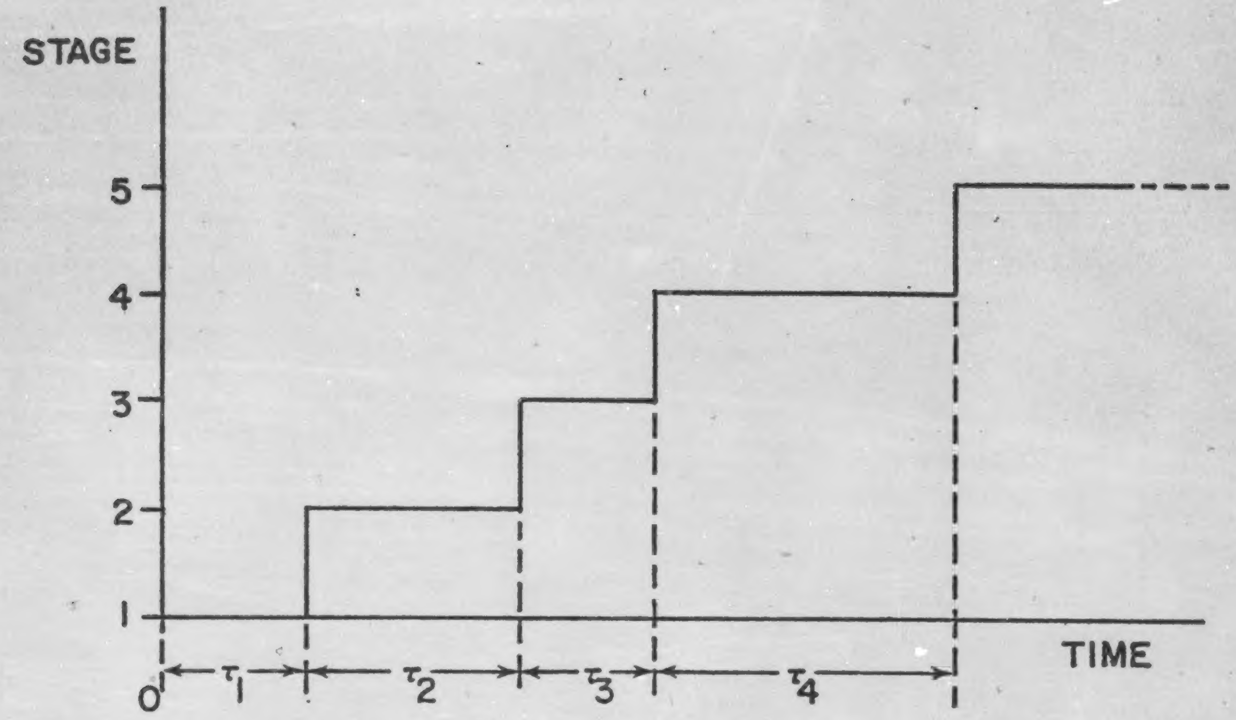


Figure 3-- A Sample Path Description Corresponding to the Unilineal Stage Structure of Figure 2.

---

[a] It is assumed that there are five stages, which must be traversed sequentially. τ_1 is the value of a random variable and denotes the sojourn time for an individual in stage i. Stage 5 is an absorbing state of the process.

$$\frac{f_1(t)dt}{1-F_i(t)} = \frac{\lambda_i e^{-\lambda_i t} dt}{1-(1-e^{-\lambda_i t})} = \lambda_i dt$$

where $f_i(t)$ is the density function corresponding to $F_i(t)$. This result, in turn, indicates that the probability of leaving stage i is independent of duration in the stage, and is tantamount to specifying an absence of aging, so new entrants have the same likelihood of departing as individuals who have been in the stage for some period of time. The parameter, λ_i, incidentally, has an interpretation as the rate of movement out of stage i; consequently, $1/\lambda_i$ equals the expected duration in stage i.

Finally, we require that

(iii) if the data pertain to the movements of a population, rather than to the transitions of a single individual, the population is homogeneous with respect to the structure of the evolutionary process.

This does not mean that all persons have the same duration τ_i in stage i, but that τ_{ic}, the time spent in stage i by individual c, follows the single exponential distribution $F_i(t) = 1-e^{-\lambda_i t}$. Stated less formally, duration in a stage is a random variable with the underlying distribution of holding times the same for all individuals. Similarly, where alternative destinations are available to persons in stage i, homogeneity means that all have the same list of probabilities for making the various transitions, not that they move identically.

It is worth dwelling on the conceptual status of the preceding assumptions. The question of the structure of M is a familiar topic to developmental psychologists, since stage theories are commonly specified at this level. Assumptions (i) to (iii) can be viewed as "side conditions," aspects of the process to which researchers have generally not been sensitive, though see Kessen (1970) and Emmerich (1968) for provocative comments on precisely these matters. What is made evident by formulating a dynamic model is that development theorists must address these auxiliary questions if complete models are to be specified. The particular assumptions we have made constitute a gross simplification of reality; this is especially true of specification (ii), which postulates an absence of duration effects, and specification (iii), which postulates population homogeneity. These assumptions do, however, provide a convenient starting point from which to consider more realistic formulations, which are developed in the next sections.

We now wish to convey the implications of assumptions (i) to (iii) for the movements of individuals among the stages. We denote by $p_{ij}(t)$ the probability that an individual in stage i at time 0 moves to stage j by time t. (This probability differs from m_{ij} in that the latter refers to movement proclivities at the occurrence of a transition, not over widely spaced time intervals.) With this specification in hand, the evolution of a population among the stages is described by the system of integral equations,

$$P_{ij}(t) = \delta_{ij} e^{-\lambda_i t} + \sum_k \int_o^t \lambda_i e^{-\lambda_i u} m_{ik} P_{kj}(t-u) du \qquad \text{(II.4)}$$

$$0 \leq i,\ j \leq n$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. This expression, known as the backward equations for a continuous-time Markov process (Feller, 1971, p. 484), is amenable to the following interpretation: (1) When $i \neq j$, $p_{ij}(t)$ consists of the sum of products of three factors: the probability of a first departure from stage i at time u, the probability of a stage i to stage k transition at that instant, and the probability of transferring to stage j by some combination of moves in the interval t - u. The summation is over all intermediate stages k and over all time divisions u in the interval (0, t). (2) When i = j, in addition to the above term, there is the possibility of not transferring out of stage i during (0, t). This probability is given by the first term.

If we represent by P(t) the matrix of elements $p_{ij}(t)$,

$$P(t) = \begin{bmatrix} p_{11}(t) & \cdots\cdots & p_{1n}(t) \\ \vdots & & \vdots \\ p_{n1}(t) & \cdots\cdots & p_{nn}(t) \end{bmatrix}$$

$0 \leq p_{ij}(t) \leq 1$, $\sum_j p_{ij}(t) = 1$, then the integral equations (II.4) have the convenient solution,

$$P(t) = e^{\Lambda[M-I]t}, \quad P(0) = I \qquad \text{(II.5)}$$

In this representation Λ is a diagonal matrix,

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

whose entries are the reciprocals of the expected duration times in each stage, I is the identity matrix, and M is the array specified in equation (II.1) which describes the pattern of movement between the stages. Further, by the expression e^A, A an arbitrary square matrix, we mean the power series in A,

$$e^{A} = \sum_{n=0}^{\infty} \frac{A^{n}}{n!} \qquad \text{(II.6)}$$

which can be evaluated by standard-numerical methods (see e.g. Gantmacher, 1960).

It is useful to recapitulate what is accomplished by this mathematical formulation. The matrix P(t) relates the distribution of a population among stages at time t to its distribution at time 0, in the sense that a typical entry, $p_{ij}(t)$, conveys the probability of moving from stage i to stage j during the interval (0,t). The model is "dynamic" in that P(t) is a function of time; with the passage of time, P(t) describes the evolution of the population among the stages. Equation (II.5) shows how the matrix P(t) is built up from the arrays M and Λ. However, while this equation is useful as a calculating formula, the logic of the process is conveyed more adequately by the integral equations (II.4).

To illustrate this model in the setting of a simple unilineal progression (matrix M_1 of Figure 2), we must specify average waiting times in stages 1,2,3, and 4. We assume these to be .5, 1, 2, and 5 years, respectively. Consequently, we have for matrix Λ,

$$\Lambda = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 \end{bmatrix} \qquad \text{(II.7)}$$

where the choice of λ_5 is arbitrary. Since stage 5 is an absorbing state, the notion of waiting time to a departure has no meaning. Mathematically, $[M - I]_{55} = [m_{55} - I] = [1-1] = 0$, so λ_5 bears no influence on the calculations.) Now, from M_1, Λ, and I, we have

$$\Lambda(M_1 - I) = \begin{bmatrix} -2 & 2 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -.5 & .5 & 0 \\ 0 & 0 & 0 & -.2 & .2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{(II.8)}$$

For the illustrative times t = 1, 2, and 4 years, we obtain, from (II.5) for P(t),

$$P(1) = \begin{bmatrix} .1353 & .4651 & .3263 & .0691 & .0041 \\ .0000 & .3679 & .4773 & .1438 & .0110 \\ .0000 & .0000 & .6065 & .3537 & .0398 \\ .0000 & .0000 & .0000 & .8187 & .1813 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix} \tag{II.9}$$

$$P(2) = \begin{bmatrix} .0183 & .2340 & .4641 & .2482 & .0354 \\ .0000 & .1353 & .4651 & .3394 & .0602 \\ .0000 & .0000 & .3679 & .5041 & .1281 \\ .0000 & .0000 & .0000 & .6703 & .3297 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix} \tag{II.10}$$

and

$$P(4) = \begin{bmatrix} .0003 & .0360 & .2881 & .4843 & .1913 \\ .0000 & .0183 & .2340 & .5079 & .2398 \\ .0000 & .0000 & .1353 & .5233 & .3413 \\ .0000 & .0000 & .0000 & .4493 & .5507 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix} \tag{II.11}$$

These values of P(t) describe the evolution of individuals among the stages, subject to the assumptions about the process structure detailed above. The entries $p_{ij}(t)$ refer to proportions[4] of the population who have moved between particular stages in the relevant time interval. For example, according to the entries in the top row of P(1), if observations are taken one year apart, we would expect 13 percent of the population in stage 1 at time 0 to still be there, 46 percent to have moved to stage 2, and 33 percent to have reached stage 3. By comparison, over a four-year interval, less than 1 percent would remain in stage 1, 48 percent would have reached stage 4, and 19 percent would be in the terminal stage of the process.

The results from the three calculations reveal that, even though the progression is unilineal with all individuals characterized by the same parameters, if observations were taken on the population at two time points, t=0 and $t=t_1$, the array[5] $\hat{P}(t_1)$ might be interpreted as evidence for a more complex theory, such as one permitting stage skipping or population heterogeneity in the rate or pattern of movement. Further, the correspondence between the matrix constructed from the population locations at two time points, $\hat{P}(t_1)$, and the rule governing stage transitions, M_1, decreases with time. Thus, different researchers observing the same population at two time points, but with different spacing intervals, might draw contrary conclusions about the stage linkage structure even though the single mechanism, M_1 of Figure 2, governs its evolution. Only with a formal model of the process could one hope to uncover its underlying structure.

III. MODELS OF MORE ELABORATE STAGE THEORIES

The matrix M contains structural information about stage linkages. Since theories of development are commonly posed at the level of specifying this array, flexibility in incorporating a variety of particular formulations would appear to be an important feature of a general framework for describing evolutionary behavior. In this section we focus on the issue of translating stage theories into M-matrices, and illustrate the evolution of P(t), the transition matrix for a population based on its locations at times 0 and t, under alternative specifications of M. As we have noted, auxiliary information about the process, concerning the distribution of waiting time intervals and the form of population heterogeneity, is required for a full description of a dynamic model. In the next section we therefore elaborate upon these "side conditions" and outline ways in which our initial assumptions can be relaxed.

No technical difficulties arise in reformulating the continuous-time Markov model to accommodate more elaborate theories of stage linkages than the structure in Figure 2. We illustrate the procedure with a few examples[6].

2. A unilineal progression which permits stage skipping. The formulation of such a structure is diagrammed in Figure 4, panel A; its translation into an M-matrix is reported in panel B. The principal new feature is that, supplementing the deterministic sequence of Figure 2, it is now possible to move

directly from stage 2 to stage 4 and from stage 3 to stage 5, when transition out of the relevant origin location takes place. We must also specify the probabilities of following the alternate paths. In the present example, lacking information as to the relative magnitudes of the various probabilities, we assume all destinations to be equally likely; that is, we prescribe $m_{23} = m_{24} = .5$, and $m_{34} = m_{35} = .5$. In practice, estimates of the transition probabilities would be assigned on the basis of theory or from observation on the empirical process.

Figure 4 about here

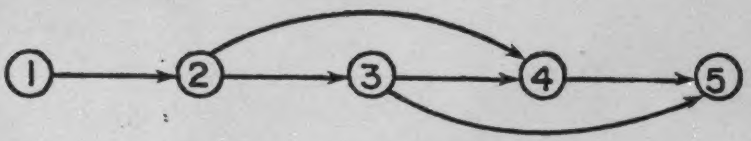
Using matrix M_2, together with the Λ array of equation (II.7), whose entries describe the rate of movement by individuals out of each stage, we obtain for P(1) and P(4), from equation (II.5):

$$P(1) = \begin{bmatrix} .1353 & .4651 & .1632 & .2012 & .0352 \\ .0000 & .3679 & .2387 & .3177 & .0757 \\ .0000 & .0000 & .6065 & .1768 & .2166 \\ .0000 & .0000 & .0000 & .8187 & .1813 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix} \qquad \text{(III.1)}$$

$$P(4) = \begin{bmatrix} .0003 & .0360 & .1440 & .4104 & .4093 \\ .0000 & .0183 & .1170 & .3964 & .4683 \\ .0000 & .0000 & .1353 & .2617 & .6030 \\ .0000 & .0000 & .0000 & .4493 & .5507 \\ .0000 & .0000 & .0000 & .0000 & 1.0000 \end{bmatrix} \qquad \text{(III.2)}$$

These P(t) arrays are the transition matrices a researcher should expect to observe if the stage locations of individuals are surveyed one year or four years apart, assuming that the population evolves according to the linkage specification M_2 together with the auxiliary conditions outlined in the preceding section. The entries are different from those obtained with the simple unilineal progression (equations II.9 and II.11), yet the same pattern of zero's

a. Diagram of Stage Linkages



b. Matrix Representation of the Stage Linkages[a]

$$M_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 \\ 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 4-- Representation of a Unilineal Progression in which Stage Skipping is Permitted.

---

[a] All destination stages corresponding to an origin location are assumed to occur with equal probability. See notes to Figure 2 for additional details on interpretation of M_2.

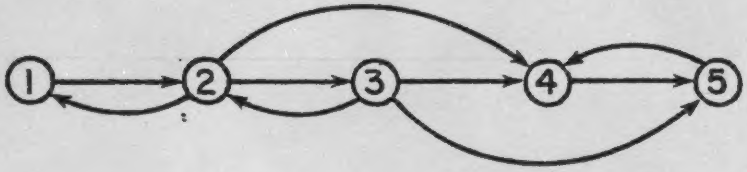
and non-zero's is present, and without a formal model of the evolution of the process a researcher would be unable to predict the different implications of these structures.

3. A unilineal progression with stage skipping and the possibility of regression. We now superimpose on the linkage structure the possibility of reverting to an earlier stage. This arrangement is diagrammed in Figure 5, panel A, in which we have provided for the possibility of backward flows from stage 2 to stage 1, from stage 3 to stage 2, and from stage 5 to stage 4. The M-matrix corresponding to this model is reported in panel B. Again, where multiple destinations correspond to an origin stage, we have arbitrarily assigned equal values to the m_{ij}'s. There is one additional alteration in M_3, in comparison with the M-matrices of earlier examples. Because there now exists a possibility of regressing from the terminal stage to an earlier level, $m_{55} \neq 1$. To maintain our conceptual imagery, in which within-stage transitions are undefined, we set $m_{54} = 1$ and $m_{55} = 0$. Note that the former value does not imply a high rate of departure from stage 5, since the rate of movement is controlled by λ_5. It only means that all transitions from stage 5 are directed to stage 4.

Figure 5 about here

To obtain P(t) we use M_3 and Λ in conjunction with equation (II.5). Here the element λ_5 in equation (II.7) is no longer arbitrary, as movement out of stage 5 is a possibility. We shall assume that such reversions are rare, and hence specify the average waiting time to a transition from stage 5 to be eight years; that is, $\lambda_5 = .125$. With these assumptions, we obtain for our illustrative calculations at t = 1, 4:

a. Diagram of Stage Linkages



b. Matrix Representation of the Stage Linkages[a]

$$M_3 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ .33 & 0 & .34 & .33 & 0 \\ 0 & .33 & 0 & .34 & .33 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 5-- Representation of a Unilineal Progression in which Stage Skipping and Regression to an Earlier Level are Permitted.

---

[a] All destination stages corresponding to an origin location are assumed to occur with equal probability. See notes to Figure 2 for additional details on interpretation of M_3.

$$P(1) = \begin{bmatrix} .2043 & .5240 & .1153 & .1374 & .0190 \\ .0871 & .4758 & .1742 & .2217 & .0411 \\ .0094 & .0858 & .6215 & .1461 & .1371 \\ .0000 & .0000 & .0000 & .8292 & .1708 \\ .0000 & .0000 & .0000 & .1067 & .8933 \end{bmatrix} \qquad (III.3)$$

$$P(4) = \begin{bmatrix} .0330 & .1560 & .1652 & .4025 & .2433 \\ .0259 & .1246 & .1500 & .4174 & .2820 \\ .0135 & .0739 & .1846 & .3542 & .3738 \\ .0000 & .0000 & 0000 & .5523 & .4477 \\ .0000 & .0000 & .0000 & .2798 & .7202 \end{bmatrix} \qquad (III.4)$$

If we compare the P(1) matrices and the P(4) matrices from the three examples [i.e., equations (II.9), (III.1), (III.3), and (II.11), (III.2), and (III.4)], we can acquire a fair idea of the implications of different stage interconnections for the evolution of a population among the statuses. We also emphasize the fact that if a population were surveyed at two time points, especially widely spaced time points, it may not be obvious from inspecting the empirically determined transition array, $\hat{P}(t_1)$, as to the structure of the stage linkages (matrix M) which generated the observations. We will return to the issue of identifying the correct structure and recovering matrix M when the observations on a process are widely spaced; first we conclude this discussion on translating theoretical specifications of stage linkages into M-matrices with a couple of examples of multilineal sequences that have been described in the developmental psychology literature.

4. _A divergent multiple progression_ (Van Den Daele, 1969, Figures 2, 4). This stage linkage structure has the diagrammatic representation of Figure 6, panel A; its corresponding M-matrix is presented in panel B. Because stages 4-7 are specified to be terminal states of the process, the corresponding rows of M_4 have 1's in the main diagonal. Van den Daele provides no discussion of waiting time distributions to departure from the various stages; hence the model remains incomplete as an evolutionary process.

Figure 6 about here

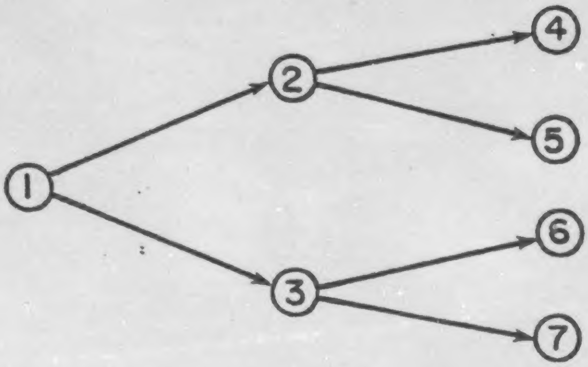
----

5. A convergent multiple progression (Van Den Daele, 1969, Figure 2). This stage sequence is depicted in Figure 7, panel A, and its associated M-matrix is reported in panel B. In this instance, the structure consists of a collection of deterministic unilineal progressions, the specific sequence for an individual being contingent upon his entry stage. Note also that the assumption of irrelevance of past history, which is posited in this formulation, is one of the side conditions we have required (assumption [i] in the preceding section). In particular, this specification appears in the fact that knowledge of the path by which one has reached stage 5 (or stage 6) is of no value in forecasting, or understanding, an individual's subsequent movements. Van Den Daele (1969) discusses several additional models of stage linkages, such as "partially convergent, divergent progression," and "partially divergent, convergent progression." As the procedure in converting flow structures into M-matrices should be evident at this point, discussions of these specifications are not presented.

Figure 7 about here

To recapitulate, subject to several side conditions, we have shown that it is possible to construct formulations of a range of developmental phenomena which mimic the evolutionary character of the observed process. With such a model one can forecast the movements of a population among the stages. By carrying out the requisite calculations for different specifications of the stage linkages, and comparing the predictions, it is possible to ascertain the

a. Diagram of Stage Linkages



b. Matrix Representation of the Stage Linkages

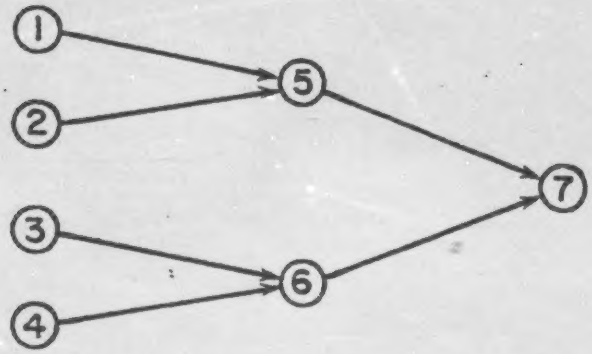
$$M_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 6-- Representation of a Divergent Multiple Progression[a].

---

[a] Source: Van Den Daele (1969, Figures 2, 3).

a. Diagram of Stage Linkages



b. Matrix Representation of the Stage Linkages[a]

$$M_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 7-- Representation of a Convergent Multiple Progression[a].

---

[a] Source: Van Den Daele (1969, Figure 2).

ways in which rather complex theories produce divergent implications and design testing schemes which maximize the possibility of rejecting one or another formulation as a description of the empirical process. Of equal importance, it is often possible to work backwards, starting with observations on the stage locations of a population at a few widely spaced time points, and derive the structure of the stage linkages compatible with the data.

An inverse problem. Until this point we have assumed that observations have been made on an empirical process in a way such that M and Λ can be estimated directly from the data, or that theories are available which specify the values of their entries. We then sought to derive the evolution of the process subject to the presumed structure. In developmental psychology, it is not uncommon for a researcher to have many observations on a few individuals (e.g., Piaget, 1954). Such a data collection scheme approximates "sample path information," a complete history on movements and waiting times of the sort illustrated in Figure 3. Detailed observations on a few subjects is a research strategy not without its costs, however. One learns little about the frequency of rare events (e.g., regression to an earlier stage, stage skipping, rare development paths) and acquires only the most rudimentary knowledge about the variation of duration times in a stage. It is therefore not surprising that investigators who rely on this approach tend to be oriented to uncovering universal rules (e.g., Piaget, 1960) rather than to elucidating individual differences and ascertaining the variety of developmental patterns.

Partly because of the limitations of small data sets, it is becoming increasingly common to employ survey methods, in which a large population, sometimes thousands of individuals, is observed (or interrogated) at a very few time points (e.g., Baltes & Nesselroade, 1972). The spacing intervals in such panel studies are usually wide, often one or more years elapse between interviews, so it is not unusual for some subjects to have made multiple moves while others have made one or zero shifts between stages. The transition matrices which can be constructed directly from such observations are P(t)-arrays, rather than M-arrays, and the stage linkages may not be readily discernible. Indeed, determination of the movement structure which underlies the evolution of the population can be a difficult task.

One approach to ascertaining the stage linkages from survey data involves consideration of the "inverse problem" to the mathematical formulation of the evolutionary model (equation II.5). Stated formally, we have available the matrix $\hat{P}(t_1)$, constructed from observations on the stage locations of individuals at times 0 and t_1. The typical entry in this matrix is $\hat{p}_{ij}(t_1) = n_{ij}(t_1)/n_{i.}$, where $n_{i.}$ = {number of individuals in stage i at time 0} and $n_{ij}(t_1)$ = {number of persons who started in stage i at time 0 and are in stage j at time t_1}. We wish to inquire whether it is possible to recover a unique M-matrix for the process and, where the answer is affirmative, we wish to estimate this matrix.

The first step in solving the inverse problem is to take the logarithm of both sides of equation (II.5).

$$Q = \Lambda[M-I] = \frac{1}{t_1} \ln \hat{P}(t_1) \qquad \text{(III.5)}$$

Just what we mean by the logarithm of matrix $\hat{P}(t_1)$, the conditions under which a solution to equation (III.5) will exist, and the circumstances under which the solution will be unique, are complex issues which are discussed at length in Singer and Spilerman (1976). Assuming we can obtain a valid and unique Q-matrix from these calculations, a second task, separating M from Λ, still remains. In many instances, though, this matter is of little concern, since the pattern of zeros and non-zeros in Q and M - I will be identical and development theories are often posed at the level of identifying permissible transitions. Moreover, because zeros are typically present in many main diagonal cells of M in models of developmental structures, a complete or near complete separation between M and Λ can frequently be effected.

We conclude this section with an example of the calculations associated with the inverse problem. Suppose observations taken on a population at times 0 and t_1 have produced the transition matrix,

$$\hat{P}(t_1) = \begin{bmatrix} .0224 & .2633 & .2402 & .1261 & .3479 \\ .0063 & .1758 & .2460 & .1735 & .3983 \\ .0216 & .0288 & .3758 & .5060 & .0679 \\ .0365 & .0745 & .0288 & .6794 & .1809 \\ .0005 & .0960 & .0460 & .0177 & .8397 \end{bmatrix} \qquad \text{(III.6)}$$

Such data would appear to be consistent with a variety of evolutionary mechanisms. From inspection of $\hat{P}(t_1)$ we do know that regression to some earlier stage must be possible; otherwise all entries below the main diagonal would be zero. Little else about the structure of M, however, can be inferred from inspection of $\hat{P}(t_1)$. Indeed, because of the sizable non-zero elements in most cells of the matrix, a researcher might conclude that direct transitions are possible between most pairs of stages.

If we are willing to assume that matrix $\hat{P}(t_1)$ was generated by a continuous-time Markov process; that is, via the evolution of the structure $P(t) = e^{\Lambda(M-I)t}$, for some matrices Λ and M which satisfy the definitional restrictions enumerated in connection with equations (II.1) and (II.5), we can solve for $\Lambda(M-I)t_1$ using equation (III.5). This yields the array,

$$\Lambda(M - I)t_1 \quad = \begin{bmatrix} -4. & 4. & 0. & 0. & 0. \\ 0. & -2. & 1. & 0. & 1. \\ 0. & 0. & -1. & 1. & 0. \\ 0.20 & 0. & 0. & -0.40 & 0.20 \\ 0. & 0.25 & 0. & 0. & -0.25 \end{bmatrix} \qquad \text{(III.7)}$$

In this instance Λt_1 and M can be separated by employing the following argument. From our earlier examples we know that a main diagonal element m_{ii} of M will equal zero if any off diagonal entry in the same row, m_{ij}, is different from zero. According to equation (III.7), each row of matrix M must have at least one non-zero off diagonal element; therefore $m_{ii} = 0$ for all values of i. With this information we can obtain Λt_1 uniquely,

$$\Lambda t_1 \quad = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & .4 & 0 \\ 0 & 0 & 0 & 0 & .25 \end{bmatrix} \qquad \text{(III.8)}$$

and solving for M provides the structure M_4 reported in Figure 8, panel A. The schematic representation of the stage linkages implied by M_4 is shown in panel B, in which probabilities of the various moves have been appended to the paths.

Figure 8 about here

The point to be emphasized is that it is not apparent from inspecting matrix $\hat{P}(t_1)$ in equation (III.6) that the underlying stage linkages are those reported in Figure 8, nor would any static analytic procedure be likely to lead a researcher to the correct conclusion. What is necessary is to construct a model of the evolution of the process and solve the implied inverse problem for the parameters which correspond to the particular data set. (In the present

a. M-matrix[b]

$$M_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 1 & 0 \\ .5 & 0 & 0 & 0 & .5 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

b. Diagram of Stage Linkages[c]

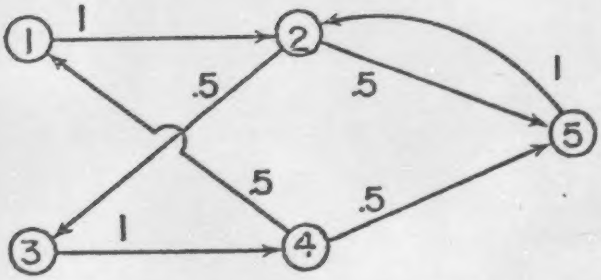


Figure 8. Stage Sequence Structure Implied by $\hat{P}(t_1)$ in Equation (III.6)[a]

example, we have assumed that the underlying model is a continuous-time Markov process [i.e., specification (i)-(iii) of the preceding section] and have solved for the matrices Λt_1 and M which are compatible with the observed array $\hat{P}(t_1)$, in that they would have given rise to this array if the postulated evolutionary process were approximately correct.)

---

[a] The process is assumed to evolve according to a continuous-time Markov formulation.

[b] Entries indicate the probability of a stage i to stage j move when transition takes place

[c] Probabilities of the various transitions are attached to the appropriate paths.

IV. ALTERNATIVE SPECIFICATIONS OF THE SIDE CONDITIONS

In this section we discuss relaxing two of the more burdensome specifications of the model, in the sense that they are likely to be inappropriate as characterizations of developmental processes. We first consider the requirement that the duration intervals in a stage must follow an exponential distribution [assumption (ii) of section II]. Following these comments we turn to the requirement that the population be homogeneous with respect to the process parameters Λ and M [assumption (iii)].

More general waiting times than exponential. The exponential distribution is frequently employed in the literature of reliability theory to describe duration intervals in a system state (stage in the current application). It has the advantages of being mathematically tractable and approximating reality in situations where the probability of a state change is uninfluenced by aging or time in the state. For example, if the process states are "alive" and "not alive," then over the middle age ranges of many animal species, the age-specific mortality rate is relatively constant and the duration intervals (in the "alive" state) are reasonably well captured by the exponential distribution. Similarly, when mortality results from exogenous events--accidents--the distribution of ages at failure can often be approximated by the exponential.

In a great many situations in social research, however, we know that proneness to changing state is a function of duration. In particular, this has been suggested with respect to residence location (McGinnis, 1968) and employment affiliation (Ginsberg, 1971). In these applications it has been argued that the duration-specific departure rate *decreases* with time, giving rise to the phenomenon of "cumulative inertia"--the longer an individual remains in a state the less likely he is to leave in the immediate future. The substantive explanations for a declining departure rate involve the growing investment an individual has made, with duration, in friendships (in the first instance) and in seniority in his place of work (in the second). There is no mathematical reason, however, to assume a declining departure rate in choosing $F_i(t)$; and in other substantive contexts a different specification may be more appropriate. For a superb review of stochastic models incorporating the notion of duration dependence, see Hoem, 1972.

A convenient way to generalize the Markov model to accommodate a variety of duration-time distributions is to begin with the integral equation representation for transition probabilities. Equation (II.4) is a special case of the formulation,

$$P_{ij}(t) = \delta_{ij}[1-F_i(t)] + \sum_k \int_0^t f_i(u) m_{ik} p_{kj}(t-u)\,du \tag{IV.1}$$

$$0 \leq i, j \leq n$$

in which the terms are identical with those of the earlier equation except that $f_i(u)$ replaces the exponential density, $\lambda_i e^{-\lambda_i u}$, and $F_i(t)$ [the distribution function corresponding to $f_i(t)$], replaces $[1-e^{-\lambda_i t}]$. A theoretically appropriate choice may now be made for $F_i(t)$.

As an illustration, one candidate for $F_i(t)$, in the case of a declining departure rate, is the two-parameter family of functions

$$F_1(t) = 1 - e^{-\lambda_i t^{\gamma_i}} \qquad \lambda_i>0 \ ; \ 0<\gamma_i<1 \tag{IV.2}$$

Here the probability of departing from state i during the infinitesimal interval (t, t + dt), conditional on the process being in state i at time t, equals

$$\frac{f_i(t)dt}{1-F_i(t)} = \frac{(\lambda_i\gamma_i t^{\gamma_i-1})\, e^{-\lambda_i t^{\gamma_i}}}{e^{-\lambda_i t^{\gamma_i}}}\, dt = \lambda_i\gamma_i t^{\gamma_i-1} dt$$

Because of the restriction on γ_i in equation IV.2, t^{γ_i-1} is a decreasing function of time, and the declining failure rate aspect of the distribution is evident.

The general formulation (IV.1) for duration time distributions and transitions between states generates a class of models known as semi-Markov processes. These generally do not have simple representations for the matrices P(t) analogous to equation (II.5), and the solution of the system of equations (IV.1) requires numerical integration methods.

Population heterogeneity. To this point we have assumed that the matrices Λ and M of equation (II.5) are identical for all individuals. This does not mean that all persons move identically since the process is probabilistic; it does imply, though, that individual level characteristics are unrelated to the structural parameters of the process. In other words, homogeneity means that considerations of genetic makeup, intelligence, sensory stimulation, and other factors by which individuals differ from one another do not portend distinct evolutionary paths in the developmental process under consideration.

There is reason to believe, however, that individual differences are present in the course of development in many processes (Werner, 1957; Kohlberg, 1968, p. 1024). We therefore desire a formulation in which the movement pattern is parametrized in terms of variables which differentiate among persons. To construct a general specification of heterogeneity within the conceptual framework of a Markov process, we assume that, corresponding to equation (II.5), the stage transitions by individual c have the structure

$$P_c(t) = e^{\Lambda_c(M_c - I)t} \tag{IV.3}$$

This formula indicates that each person is characterized by a pair of matrices, Λ_c and M_c, and his evolution, in turn, is described by $P_c(t)$. Thus, our formulation begins with a separate Markov process for each individual.

This approach directs a researcher to identify the variables which describe heterogeneity; that is, to ascertain which factors account for individual differences in the matrices M and Λ. Thus, not only does a heterogeneity formulation lead to more realistic models of evolutionary processes, in that allowance is made for individual differences, but it stresses the analytic tasks of specifying the variety of developmental patterns in a population and ascertaining the attributes which make an individual more prone to following one set of paths rather than another.

One form of heterogeneity concerns the distribution of M-matrices in a population. Focusing on these arrays serves to emphasize individual differences in proneness to making particular moves where a transition takes place. We shall not discuss this form of heterogeneity in the present essay and direct the interested reader instead to McFarland (1970), Spilerman (1972a) and Singer and Spilerman (1974). A second form of heterogeneity stresses individual differences in the Λ-matrix, i.e., in the rates at which departures occur for persons in the various states. We conclude this section with a simple formulation of population heterogeneity in which it is assumed that the individual differences can be expressed in the latter way.

To simplify the discussion, we further require the non-zero entries in the diagonal matrix Λ to be equal for an individual; i.e., $\lambda_i = \lambda$ for all i. This means we are specifying identical departure rates from all states. As a result, equation (II.5) reduces to

$$P(t|\lambda) = e^{\lambda t(M-I)} \qquad \text{(IV.4)}$$

where $P(t|\lambda)$ denotes the transition matrix for an individual having a rate of movement value equal to λ. We shall assume that equation (IV.4) describes the evolution of an individual drawn at random from the population.

Heterogeneity is incorporated into the formulation by specifying a density function $g(\lambda)$ which describes the distribution of λ-values in the population. We now define the population-level transition matrix corresponding to times 0 and t to be

$$P(t) = \int_0^\infty P(t|\lambda)g(\lambda)d\lambda = \int_0^\infty e^{t\lambda(M-I)}\, g(\lambda)d\lambda \qquad \text{(IV.5)}$$

This formula expresses the population-level matrix as a weighted average of the individual-level arrays, $P(t|\lambda)$, the weights reflecting the population proportions associated with particular λ-values.

To complete this specification of heterogeneity it is necessary to select a density function $g(\lambda)$ to describe the distribution of λ-values. One useful choice is the gamma family of functions

$$g(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \qquad \lambda>0 \text{ , } \alpha>0 \text{ , } \beta>0 \qquad \text{(IV.6)}$$

which is flexible enough to describe a variety of unimodal curves. With this selection of $g(\lambda)$, a convenient representation of the population-level matrix $P(t)$ is obtained (Spilerman, 1972b, p. 608):

$$P(t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha} \left[I - \frac{t}{\beta+t} M\right]^{-\alpha} \qquad \text{(IV.7)}$$

The transition probabilities (IV.7) do not describe the evolution of a Markov process; however they do describe the movement of a population *in which each individual follows a Markov model* with individual differences being specified by $g(\lambda)$ in equation (IV.6).

In analogy with our earlier inverse problem discussion for Markov chains, the present formulation can be used with observations taken at widely spaced time points, 0 and t_1, together with estimates of α and β to yield an estimate of the underlying transition mechanism M, according to the matrix equation

$$M = \left(\frac{\beta+t_1}{t_1}\right) \left[I - \frac{\beta}{\beta+t_1} \left(\hat{P}(t_1)\right)^{-\frac{1}{\alpha}}\right] \qquad \text{(IV.8)}$$

Thus, from observations of the sort collected in many surveys, even under an assumption of population heterogeneity in the rate of movement, it may be possible to recover the matrix of stage linkages which governs the evolution of the process.

V. CONCLUSIONS AND SUMMARY

In this paper we have explored the consequences of particular stage linkage structures for the evolution of a population. One thrust of our comments has been to identify the sorts of process features concerning which assumptions must be made in order to convert a static theory about stage connections into a dynamic model. A second focus in our discussion has centered on inverse problems; how to utilize a model formulation so that the stage linkage structure (matrix M) may be recovered from survey data of the kind usually collected by developmental psychologists.

We have presented only the most rudimentary sorts of stage structures. Indeed, even within the Markov framework we have limited our consideration to a subset of these models; namely, those which are time-stationary (i.e., Λ and M are not functions of time). By this specification we have excluded the possibility of accomodating age-dependent transition laws, a consideration of substantial importance in developmental psychology. (An extension of the models discussed here to incorporate both age dependence and cohort effects is, however, a feasible undertaking but one with an increase in mathematical complexity.) Further, all the models we have discussed entail a low dependence of future movements on the transition history of an individual, given his current stage[7]. Restrictions of these sorts are likely to be reasonable for some processes, unreasonable for others. Appropriate models of developmental phenomena must, therefore, be constructed from a list of known characteristics about an empirical process.

We also point out that the concept of stage merges with the notion of state as the number and sorts of permissible transitions increase. "Stage" seems conceptually rooted to the idea of progress (i.e., development) and would be an appropriate component of a theory which sees the system's statuses as genetically determined or as facilitating the conditions for succeeding statuses to come into play.[8] The mathematical framework we have introduced is also compatible with a "state" notion, in which there is an extensive opportunity to cycle among the statuses. State formulations have been suggested in the psychology literature in relation to anxiety, moods, etc. (e.g., Kessen, 1962, pp. 72-73).

As a final set of considerations in relation to the structure of stage models, we note that all the formulations we have addressed are models of solitary processes. We have proceeded as if intelligence, cognition, motor skills, and personality development unfold autonomously. In reality there no doubt exist extensive dependencies among some of these processes. Mathematical models of interacting developmental phenomena could be formulated but clear empirically-based specifications of such dependencies are still lacking.

FOOTNOTES

[1] The initial condition, $y(0) = 1$, in equation (2) is necessary because diffusion through communication cannot begin until at least one person is knowledgeable.

[2] For a more technical presentation of continuous-time Markov processes see Feller (1968, Chap. 17) and Singer and Spilerman (1974). For discussions on the superimposition of theoretical structures on stochastic models see Coleman (1964, Chaps. 5, 6).

[3] In the present example $i, j, k = 1, 2, 3$, respectively.

[4] If the observations are on a single individual the interpretation of $p_{ij}(t)$ is in terms of the probability of a stage i to stage j move between times 0 and t.

[5] The symbol "^" over a matrix or over an element in a matrix, will mean that it should be viewed as estimated directly from data rather than calculated from a mathematical model.

[6] We begin here with example 2; example 1 refers to the structure in Figure 2.

[7] The time-stationary Markov formulations postulate irrelevance of prior stage affiliations, durations in those stages, and duration in current stage. The last two of these restrictions can be eliminated by introducing non-stationary semi-Markov models as delineated, for example, in Hoem, 1972.

[8] Stages in childhood, such as "walking" or "reading" expose an individual to entirely new sets of experiences which may be prerequisites for the onset of more advanced behaviors.

Mathematical Description and Repesentation of Developmental Change Functions on the Intra- and Interindividual Levels

by

Kenneth E. Guire
Center for Human Growth
and Development

Charles J. Kowalski
Dental Research Institute

The University of Michigan

Table of Contents

MATHEMATICAL DESCRIPTION AND REPRESENTATION OF DEVELOPMENTAL CHANGE FUNCTIONS ON THE INTRA- AND INTERINDIVIDUAL LEVELS

Kenneth E. Guire
Center for Human Growth
and Development

Charles J. Kowalski
Dental Research Institute

The University of Michigan

ABSTRACT
I. INTRODUCTION
II. DESIGNS FOR THE STUDY OF DEVELOPMENT
III. MODELS FOR INTRAINDIVIDUAL CHANGE
IV. MODELS FOR INTERINDIVIDUAL CHANGE
V. SUMMARY

ABSTRACT

Methods of the analysis of longitudinal data are discussed. For the intra-individual case which deals with a single longitudinal series, a variety of curve fitting approaches are considered. For the inter-individual case which deals with one or more samples of longitudinal observations, a variety of methods including univariate analysis of variance, multivariate analysis of variance, and polynomial growth curve models are considered. An attempt is made to relate these methods, which are widely used in biological applications, to the broader context of developmental research.

Mathematical Description and Representation of Developmental Change Functions on the Intra- and Interindividual Levels

I. INTRODUCTION

In an earlier paper (Kowalski & Guire, 1974), we surveyed the then available data analytic strategies for several types of longitudinal data sets. In particular, we identified six distinct types of longitudinal data sets, viz., (1) univariate time series, (2) univariate one-sample data matrices, (3) univariate K-sample data matrices, (4) multivariate time series, (5) multivariate one-sample data matrices, and (6) multivariate K-sample data matrices, and the methodologies appropriate for each of these types were treated separately. The present paper builds on this background, paying special attention to the mathematical description and representation of developmental change functions on the intra- and interindividual levels. An overview of the available models and statistical procedures for the analysis of such data is presented. The aim of this presentation is to identify certain procedures which have proved to be useful in the biological, growth-oriented sciences for possible application in behavioral developmental research. In so doing, we must carefully consider potential differences in the structure of the measurements to be analyzed and, whenever possible, models which do not require monotonicity and nonparametric analogs for the statistical procedures discussed will be cited. On the other hand, more research - both from the standpoint of theory and from the standpoint of practice - needs to be done before we can confidently analyze multivariate data sets in any field of application (c.f. Kowalski, 1972), so that any words of caution in this context should be tempered by the need for the entire developmental research community to gain some experience in the use of these techniques (Prahl-Andersen & Kowalski, 1973). It is in this spirit that this paper is written. Behavioral scientists should be aware of the possibilities and limitations of statistical treatment of developmental data and the models presented were chosen to illustrate both ends of this spectrum, as well as many of the shady areas in-between.

We begin with a general discussion of longitudinal data sets in the context of the design of developmental studies. The purpose of this section is to place the longitudinal approach into some perspective vis-a-vis the oft heard dictum that "the design of a study is a function of the purposes of the investigation" and an attempt at delineating the class of studies for which longitudinal designs may be appropriate is made. It is recognized that development will often depend on factors other than simple chronological age, but it is argued that this does not imply that the class referred to above is empty. Since questions of this type are considered in great detail in the first part of this book, our discussion is brief and somewhat cursory. It is included only in an attempt to counteract some of the impact of much of the recent developmental literature which seems bent on condemning the longitudinal method. We then consider questions associated with the descriptive and explanatory study of intraindividual change. We focus on the implications of choosing one or another of the models which can be used for these purposes and thereby confront important philosophical problems ranging from the making of reasonable a priori assumptions to the validation of a model by the expedient of subjecting it to a goodness-of-fit test. Finally, we discuss techniques for the investigation of interindividual differences in development. We consider not only formal hypothesis testing techniques but also less formal, descriptive, data-analytic procedures which may prove useful in unraveling some of the complex problems associated with the measurement of change (Harris, 1963).

II. DESIGNS FOR THE STUDY OF DEVELOPMENT

Much of the current thinking in behavioral research concerning the study of development was generated by a series of papers by Schaie (1965, 1970, 1972). In 1965, Schaie introduced a trifactorial developmental model which views development, D, as a function of A = chronological age, C = cohort and T = time-of measurement. Perhaps the greatest contribution of this model was to focus attention on the implications it had for the choice of the design of developmental studies. In particular, Schaie (1970) pointed out that (a) cross-sectional designs confound the age and cohort effects, (b) longitudinal designs confound the age and time-of-measurement effects and (c) time-lag designs confound the cohort and time-of-measurement effects. While these facts were apparent long before Schaie introduced his model, the model provided a convenient conceptual framework which clearly illustrated the source of these problems. In an attempt to rectify the situation, the general notion of a mixed-longitudinal design was developed (Prahl-Andersen & Kowalski, 1973) and several special cases, viz., the cohort-sequential, time-sequential and cross-sequential designs, were identified and proposed for use in certain well-defined types of developmental investigations (Wohlwill, 1970). It was clear that Schaie viewed these mixed-longitudinal strategies as completely replacing the more traditional designs. Schaie (1972) was especially vociferous in condemning longitudinal designs, concluding that, "the single cohort longitudinal study be used for no other purpose than that of the historian, the case history reporter, or to gather anecdotal material for the purpose of generating hypotheses." However, none of the Schaie designs completely solves the problem of confounding alluded to earlier and a number of questions have arisen regarding both the analysis (van't Hof, et al., 1976) and the interpretation (Baltes, 1968) of data collected using the mixed-longitudinal approach. Thus Hindley (1972) responded to Schaie by noting, "...at the same time that he belabors the longitudinal method for resting on dubious assumptions, he seems prepared to make other assumptions which some might consider equally open to question.

One example is his use of projected longitudinal gradients which, as he admits, depend on the assumption that environmental input will be approximately equal over past and future time intervals" (p.40). It would appear, then, that there may be a bit of life left in longitudinal research despite the well-known and documented problems it presents in the data collection (Jones, 1958), data-analytic (Kowalski & Guire, 1974), and interpretive (Cronbach & Furby, 1970) spheres.

We approach this question by summarizing the case against the longitudinal method in the context of Schaie's trifactorial developmental model. If we write $D = f(A,C,T)$ to represent Schaie's model, we see that there are two distinct problems to be faced. The first is that since longitudinal studies are indexed by but two of the three factors comprising Schaie's model, $D = f(A,T)$, we cannot generalize the results of any longitudinal investigation beyond the cohort selected for study. Otherwise stated, if we are to attempt to generalize the results of a longitudinal study, we must assume that the cohort effect is zero. The second problem already mentioned, is that even if the cohort effect can reasonably be neglected, age differences will still be confounded with time-of-measurement effects, these being viewed as temporary variations or aberrations superimposed on the developmental function. What Schaie's argument comes down to, then, is that longitudinal studies are appropriate only in those cases when $C = T = 0$. But must every study answer all questions? It is our opinion that to adopt this attitude would stifle a good deal of potentially valuable research. There are situations in which the longitudinal approach must be employed since a number of questions cannot be answered in any other way. To cite but one example, if our concern is with intraindividual patterning of relationships among variables over time, there simply is no substitute for the longitudinal method (Wohlwill, 1970). Whenever we wish to study the relationships between the amount or direction of change for two or more variables, or between such change and any other information about the individual, the only alternative to the longitudinal approach is the abandonment of the project. This does not mean that every project is worth doing, that one should flatly ignore potential time-of-measurement disturbances, and that cohort effects are mere figments of the imagination. It means simply that we should take care with the procedural conduct of the study and prudently limit our inferences to the population from which our sample was selected. We need not over-react to the point of jettisoning the longitudinal approach. As stated by Jones (1958), "If we wish...to achieve a body of developmental theory, we cannot eliminate developmental observation" (p.98).

While we certainly do not advocate the use of the longitudinal approach in every developmental investigation, it is our contention that there are situations in which time-of-measurement effects may be safely neglected, e.g., in most studies of physical growth, and that when these effects are unimportant, most of the objections to the longitudinal approach disappear. This is primarily due to the fact that the remaining technical difficulties associated with the longitudinal method are counterbalanced by its great efficiency in estimating change scores. Wallis and Roberts (1956), e.g., estimated that in analyzing the weights of men before and after a lapse of time, each of two independent samples would have to contain 2,222 individuals (a total of 4,444 observations) to provide the same sampling reliability as a single paired-sample of 25 men measured before and after the lapse of time. While the data they used to obtain these estimates were artificially generated, they do provide some idea of the increase in precision of the paired-sample approach which may be expected in practice (see also Rao & Rao, 1966).

We turn now to questions dealing with the study of intraindividual change. Some mathematical models which can be used to mirror these changes are presented and discussed in the context of their potential usefulness in developmental research.

III. MODELS FOR INTRAINDIVIDUAL CHANGE

We begin, following Kessen (1960), by agreeing that "a characteristic is said to be developmental if it can be related to age in an orderly or lawful way." Thus, given a series of measurements $x_1, x_2, \ldots, x_T$ on a given individual, we suppose that the t-th such measurement $(t=1,2,\ldots,T)$ can be expressed in the form

$$x_t = f(t) + \varepsilon_t \qquad \text{(III.1)}$$

in which the observations x_t are viewed as being composed of a systematic part, $f(t)$, and a random or stochastic part, ε_t, which obeys some probability law. The basic problem is then to fit a function, $f(t)$ to the observations in such a way that the function (a) provides a close fit to the data, (b) has a reasonably simple mathematical structure and (c) has relatively few parameters, whose meanings are clear with a definite developmental significance (Israelsohn, 1960). A number of such functions have been proposed for use in a variety of developmental circumstances. Perhaps the simplest of these is the first-order autoregressive scheme, or Markov Process, in which

$$f(t) = \beta x_{t-1}$$

so that the value of the observation at time t is a simple linear function of the measurement made at the preceding time point. The next most complex form of linear autoregressive series is the Yule series where

$$f(t) = \beta_1 x_{t-1} + \beta_2 x_{t-2}$$

and x_t is determined by the values of the observations made at the preceding two time points. This formulation can be extended in obvious ways and a good account of autoregressive models is provided by Kendall and Stuart (1968). These and a number of other forms for f(t) may be generated by characterizing the developmental process in terms of a differential equation and we here sketch some examples of simple differential equations which have been used to this end. Letting t denote time and x the magnitude of the measurement being taken, the differential coefficient dx/dt then denotes the rate of growth, i.e., the increase in x per unit time. It is generally assumed that the growth process may be characterized by a differential equation

$$\frac{dx}{dt} = g(x,t)$$

which says that the growth rate depends both on time and current size. In the examples to follow, we consider only special cases of the type

$$\frac{dx}{dt} = g(x)h(t) \qquad \text{(III.2)}$$

which may be written as

$$\frac{dx}{g(x)} = h(t)dt$$

or, solving,

$$\int \frac{dx}{g(x)} = H(t) \qquad \text{(III.3)}$$

which determines x as a function of t. Turning to some specific examples, if we let g(x)=1, x, λ-x and x(λ-x) for 0 < x < λ in (III.2), we obtain the differential equations

$$\frac{dx}{dt} = \begin{matrix} h(t) \\ xh(t) \\ (\lambda-x)h(t) \\ x(\lambda-x)h(t) \end{matrix} \qquad \text{(III.4)}$$

where, in the last two equations, λ is interpreted as the maximum value of x. The four equations relate to quite different and varied types of growth processes. Specifically, the respective equations indicate that at a given time the growth rate (1) depends on time, but not on size, (2) is proportional to size and a function of time, (3) is proportional to the "growth potential," i.e., the maximum size minus current size, and a function of time and (4) is proportional to both the current size and growth potential, as well as a function of time.

If we now consider the "logarithmic differential coefficient" $d\log x/dt = dx/xdt$ which denotes the relative growth rate, i.e., the proportional increase per unit of time, the last three equations in (III.4) may be written

$$\frac{d\log x}{dt} = h(t)$$

$$\frac{-d\log(\lambda - x)}{dt} = h(t)$$

and

$$\frac{d\log x}{dt} - \frac{d\log(\lambda - x)}{dt} = h(t)$$

Solving these as in (III.3), the equations (III.4) yield

$$x = H(t)$$
$$\log x = H(t)$$
$$\log(\lambda - x) = \log\lambda - H(t)$$
$$\log(\lambda - x) - \log\lambda = -\lambda H(t)$$

or, if solved for x,

$$x = \left\{ \begin{array}{l} H(t) \\ \exp(H(t)) \\ \lambda[1-\exp(-H(t))] \\ \lambda[1+\exp(-\lambda H(t))]^{-1} \end{array} \right. \qquad \text{(III.5)}$$

Here $\exp(H(t)) = e^{H(t)}$ where e is the base of the natural logarithms. The equations include a constant of integration which may be determined from a given value of (x,t). By looking at particular values of h(t) we can now generate a number of examples of growth curves satisfying the conditions set out following equation (III.4). Taking $h(t) = \beta$, e.g., we obtain

$$x = \left\{ \begin{array}{l} \alpha + \beta t \\ \exp(\alpha + \beta t) \\ \lambda[1 - \exp(-\alpha - \beta t)] \\ \lambda[1 + \exp(-\lambda(\alpha + \beta t))]^{-1} \end{array} \right.$$

For $\beta > 0$ these are increasing functions of t, the last two having asymptote λ. The last of these expressions defines what is generally called the logistic growth curve. By retracing the steps that led to its derivation, we may be able to gain some appreciation for the sorts of growth processes it might reasonably be expected to characterize. The equation for growth rate is

$$\frac{dx}{dt} = \beta x(\lambda - x)$$

and for relative growth rate,

$$\frac{d \log x}{dt} - \frac{d \log(\lambda - x)}{dt} = \lambda\beta$$

or, since

$$\frac{d \log(\lambda - x)}{dt} = \frac{1}{\lambda - x}\,\frac{dx}{dt} = -\frac{1}{\lambda - x}\,\beta x(\lambda - x) = -\beta x$$

we can write

$$\frac{d \log x}{dt} = \beta(\lambda - x)$$

i.e., the relative growth rate is a linear function of x. Thus before fitting a logistic function to developmental data, one should be sure that the conditions implied by these equations do not violate their a priori knowledge of the process under consideration. Looked at the other way around, after fitting a logistic function to developmental data, a reasonable test of goodness-of-fit would be to plot the values of x on the abscissa vs. the values of $\Delta \log x / \Delta t$ on the ordinate to see whether or not a linear relationship obtains. But it should be noted that while goodness-of-fit is perhaps a necessary condition for the employment of a particular function to mirror a growth process, it is by no means sufficient to ensure transcending mere description to the real desiderata of explanation. This is due not only to technical, statistical difficulties (Kowalski, 1970, 1972), but also to the very philosophy underlying the use of goodness-of-fit tests in this context. As stated by Feller (1966),

> The logistic distribution function...may serve as a warning. An unbelievably huge literature tried to establish a transcendental "law of logistic growth": measured in appropriate units, practically all growth processes were supposed to be represented by a function of this form... Lengthy tables, complete with chi-squared tests, supported this thesis for human populations, bacterial colonies, development of railroads, etc. Both height and weight were found to follow the logistic law even though it is theoretically clear that these two variables cannot be subject to the same distribution. Laboratory experiments on bacteria showed that not even systematic disturbances can produce other results. Population theory relied on logistic extrapolations (even though they were demonstrably unreliable). The only trouble with the theory is that not only the logistic distribution, but also the normal, the Cauchy, and other distributions can be fitted to the same material with the same or better

> goodness of fit. In this competition the logistic distribution plays no distinguished role whatever; most contradictory theoretical models can be supported by the same observational material. (p. 52)

Thus the proper emphasis on fitting a curve to longitudinal data is not on selecting a function on the basis of goodness-of-fit, but rather on selecting a function which accurately mirrors the biological structure of the process under consideration. There are certainly enough functions to choose from -- each with its own set of assumptions which must be met if we are to go beyond a mere description of our developmental data. In addition to those already discussed, we should mention several others that have been proposed for use in relatively well-defined sets of circumstances. In the realm of physical growth, because of the adolescent growth spurt typical of the higher primates (which may or may not obtain in psychosocial investigations), a parameterization consisting of distinct components for prepubertal and adolescent growth is often recommended. Thus Deming (1937) suggested the use of

$$f(t) = \alpha + \beta t + \gamma \log(t)$$

for the period up to nine years in girls and ten in boys, and from that point to maturity, the Gompertz (1825) curve, viz.,

$$f(t) = \alpha \exp[-\exp(\beta - \gamma t)]$$

Similarly, Jenss and Bayley (1937) fit

$$f(t) = \alpha + \beta t - \exp(\gamma + \delta t)$$

over the prepubertal period and then used the Gompertz function. An analogous strategy was suggested by Count (1943). Examples were provided by Israelsohn (1960). But, as pointed out by Bock et al. (1973), the problem as to where one curve should end and the other begin is still an open question. They suggested instead the use of a mixture of logistic growth curves. In their notation,

$$f(t) = \frac{a_1}{1 + \exp[-b_1(t-c_1)]} + \frac{f-a_1}{1 + \exp[-b_2(t-c_2)]}$$

where

a_1 is the upper limit of the prepubertal component.

b_1 determines the initial slope of the prepubertal component, implicitly given by $v_1 = a_1b_1/4$, the maximum velocity of growth in the prepubertal component.

c_1 determines the location in time of the prepubertal component.

f is mature size.

$a_2 = f-a_1$ is the contribution of the adolescent component to mature size.

b_2 determines the slope of the adolescent component, implicitly given by $v_2 = a_2b_2/4$, the maximum velocity of growth of the adolescent component.

c_2 is the age at maximum velocity of the adolescent component.

In fitting this model to data on stature, Bock et al. (1973) were given t, observed y, assumed f known and the remaining five parameters (a_1, b_1, c_1, b_2, c_2) were fit by non-linear least-squares. Another parameterization which permits straightforward interpretation of the parameters comprising the model was suggested by Weinbach (1941). Here

$$f(t) = b_1 \exp(c_1t) - b_2 \exp(-c_2t)$$

where

c_1 is the multiplicative rate of growth per unit time.

b_1 is the size of the individual when he enters the time span of constant multiplicative growth in early or middle childhood.

b_2 is birth weight

and

c_2 represents how rapidly the child decelerates from birth into the phase of constant multiplicative growth.

The rationale behind the use of this model is that since the growth of most physical measurements is decelerative in infancy and more nearly constant for some years thereafter, a convenient mathematical representation of the growth of an individual is one which estimates both this deceleration and the more constant phase of middle childhood. Presumably the use of another function would be required if the age range were extended to include the pubertal spurt.

We might also mention at this stage another model which can actually be used to test the hypothesis of a significant change in the pattern of growth due to some event E (e.g., puberty) occurring within the interval of observation. This is due to Box (1967) who considered the general problem of testing for a change in the level of a non-stationary time series. Potential applications in the context of the present discussion include checking on whether or not behavioral measurements exhibit a growth spurt and in facilitating the choice of where different growth curves may be needed to accurately mirror changes in the processes governing development. Suppose we have a total of $T = n + m$ measurements, the first n of these being taken before E, the next m after. If then δ measures the shift in level of the series associated with the event E, Box's model is of the form

$$x_t = \begin{cases} L + \gamma_o \sum_{j=1}^{t-1} \alpha_{t-j} + \alpha_t & \text{for } t \leq n \\ L + \delta + \gamma_o \sum_{j=1}^{t-1} \alpha_{t-j} + \alpha_t & \text{for } t > n \end{cases}$$

where L denotes the initial location of the series, γ_o is a constant, $0 \leq \gamma_o < 2$, presumed known, and the α's are independent normal deviates having variance σ^2. It may aid in the interpretation of this model to write

$$\gamma_o \sum_{j=1}^{t-1} \alpha_{t-j} + \alpha_t = \gamma_o \sum_{j=1}^{t-2} (1-\gamma_o)^j x_{t-1-j} + (1-\gamma_o)^{t-1} L + \alpha_t$$

which emphasizes its autoregressive structure. Box then shows how to estimate σ^2, L and δ (say by s^2, $\hat{L}$ and $\hat{\delta}$) from the data and the required test follows from the fact that

$$(\delta-\hat{\delta}) \left\{ \frac{[1-(1-\gamma_o)^{2n}]\,[1-(1-\gamma_o)^{2m}]}{[1-(1-\gamma_o)^{2t}]\,\gamma_o\,(2-\gamma_o)\,s^2} \right\}^{\frac{1}{2}} \qquad \text{(III.6)}$$

has Student's t-distribution with n + m - 2 degrees of freedom. Box approached this problem from the Bayesian point of view in which certain (non-informative) prior distributions for the parameters in the model were assumed, (III.6) then representing the posterior distribution of δ. The test can, however, be directly applied in the more usual Neyman-Pearson framework where no a priori information concerning these parameters is invoked. In either case, γ_o is taken as known but Box has shown that (III.6) is relatively insensitive to changes in the value of γ_o.

The point of the above examples is to acquaint the reader with a number of models which have been proposed for representing intraindividual physical growth. As already noted, it is important to realize that in the competition between these models, goodness-of-fit plays a relatively minor role. While a poor fit of the model to the data should reasonably cause one to question the applicability of the model under consideration in the context of the current problem, a good fit to the data is not sufficient to insure this applicability. A more prudent course is perhaps via the derivation of a model that satisfies certain definite a priori requirements imposed by the structure of the developmental process under consideration. This may be approached by the use of differential equations as sketched above (see also Shock, 1951). Alternatively, the properties of available models can be checked to see whether or not they conform to these a priori criteria. Thus, for example, if we wish to use a model that is consistent with allometric growth, the use of the Gompertz curve may be appropriate (Deakin, 1970).

On the other hand, if only a simple descriptive function is required and/or little is known about the mechanisms governing the growth process, the class of polynomial functions

$$f(t) = a_o + a_1 t + a_2 t^2 + \ldots + a_p t^p$$

are apt to be satisfactory and have the convenient property that the "mean curve" (that fitted to the observed growth patterns of a number of individuals) is equivalent to the "mean constant curve" (that obtained by fitting the individual records to a set of such polynomials and then averaging the coefficents a_i). This is not true for growth curves in general (e.g., Gompertz, logistic) and thus the character of the individual curves are subject to distortion through group averaging. This may be a critical point in practice since indiscriminate averaging tends to over-smooth the growth curves, masking the inherent interindividual variability, which is often of prime importance in the study of growth. Thus while polynomial growth curves may not lend themselves to easily interpretable explanatory models for growth processes, they may still be useful for the description of development and in the effective reduction of the observations to a small number of parameters characterizing the observed course of growth. This method was introduced by Wishart (1938, 1939) who suggested that the growth curve for each subject be broken down into its mean and linear, quadratic, etc. components, each of these being subjected to separate analysis. The effects of treatments on the average growth rate could then be seen from the analysis of the linear components, and analysis of the higher-order components would show to what extent the treatments were effecting the shapes of the growth curves. The method was valuable in that it succeeded in replacing the successive observations on growth by a few summary figures which led to efficient comparisons between the groups being studied (Rao, 1958).

In an attempt to extend this approach, Rao (1958) considered the problem of transforming time by a function $\tau = G(t)$ in such a way that the growth rate is uniform with respect to this new time metameter, so that an adequate representation of growth would be available in terms of the initial value of the measurement and the redefined uniform rate. This method produces the required transformation from the data in hand, provides a valid test of the hypothesis that the average growth curve is the same under all treatment conditions irrespective of any assumptions on the nature of the growth curve, and it is not even necessary to know the exact values of the time points at which the observations were made. Rao (1958) also considered the model

$$y_{t\alpha} = \lambda_{\alpha} g(t) + \varepsilon_t \qquad \text{(III.7)}$$

where $y_{t\alpha}$ is the increase in the t-th interval, λ_a is a parameter specific to individual α, $g(t)$ is an unknown function of time only, and ε_t is a random error. Whereas the first method did not depend on any assumptions about the individual growth curves, (III.7) implies that, apart from a deterministic linear trend for growth with respect to some time metameter, there are independent disturbances taking place in small intervals of time. By a common transformation $\tau = g(t)$, all the individual growth curves can be made linear apart from random fluctuations.

Finally, Rao also considered extending (III.7) to its factor-analytic analog

$$y_{t\alpha} = \lambda_{\alpha}^{(1)} g_1(t) + \lambda_{\alpha}^{(2)} g_2(t) + \ldots + \varepsilon_t \qquad \text{(III.8)}$$

where $\lambda^{(1)}$, $\lambda^{(2)}$,... correspond to the factors and g_1, g_2 ... to the regression coefficients. If (III.8) holds, we should be able to replace the growth curve by its estimated factor values $\lambda^{(1)}$, $\lambda^{(2)}$,... and to single out the dominant ones for further analysis. While this approach has obvious merit as a potentially valuable data-reduction technique, (III.8) differs enough from the standard factor analysis model to require an entirely new set of associated significance tests and these have not as yet been worked out. In case ε_t can be assumed independent of t, Hotelling's principal component analysis may be used to obtain the requisite factors and standard tests can be applied (Rao, 1958).

In the following sections we consider how some of these models for intraindividual development are used in the study of interindividual differences in developmental patterns and in providing tests of hypothesis concerning the mean patterns of growth in several groups of individuals.

IV. MODELS FOR INTERINDIVIDUAL CHANGE

When an investigator is concerned with a single attribute measured longitudinally on one or more groups of individuals, there are a variety of analytical models which can be employed. These techniques which are quite different from those described above for intraindividual analysis, fall into three main categories: (1) univariate analysis of variance, (2) multivariate analysis of variance, and (3) polynomial growth curve (PGC) models. For any particular analysis problem, the choice of one of these three approaches should be made as a function of the extent to which the structure of the model is appropriate and the extent to which the statistical assumptions are met. In describing these three approaches, careful attention will be paid to these points. However, most attention will be given to the PGC models which are least widely considered in applications.

The univariate analysis of variance models are probably the most widely used, most widely documented (viz. Winer, 1962; Gaito & Wiley, 1963), and most problematic approach to the analysis of longitudinal data. In the case of a single sample of individuals, the approach is often referred to as trend analysis (Winer, 1962; Kowalski & Guire, 1974). In this model the total sum of squares is partitioned into components attributable to individual differences, time, and error under the assumption of no interaction between the time and individual factors. This model allows the investigator to test the overall hypothesis of no differences attributable to the time factor. It is also possible to subdivide the sum of squares for time into orthogonal polynomial components allowing hypotheses concerning the shape of the time response to be tested.

In the case of two or more samples of individuals measured longitudinally, a repeated measures analysis of variance (Winer, 1962) can be employed. In this model, individuals are treated as a random factor nested within groups with repeated measurements over time. In the context of this model the main null hypotheses of interest are (1) no time effect, (2) no group effect, and (3) no time by group interaction. The last of these hypotheses is often of greatest interest since it can be thought of as a test that the time response functions of the k groups are parallel. As in the simpler case described above, it is possible to partition the time effect into orthogonal polynomial components to gain greater insight into the shape of the time response. This k-sample repeated measure design can be thought of as a prototype for a great variety of more complex models in which the k-groups are structured as the levels of a factorial or other design. It is also possible to structure the repeated measure as levels of a more complex experiment.

These analysis of variance models seem at first glance to be ideal for the analysis of longitudinal data since they are relatively simple and the questions of interest correspond to hypotheses which can be tested in the context of these models. The problem, of course, has to do with the validity of the underlying statistical assumptions of the models. It is an unfortunate fact that the ratios of mean squares will have an exact F-distribution only under rather restrictive assumptions described by Huynh and Feldt (1970). A sufficient condition for the result requires that the repeated measures are normally distributed, have equal variances, and either are mutually independent or have equal correlations (Greenhouse & Geisser, 1959). The assumption of mutual independence is virtually never tenable and the assumption of equal correlations is seldom tenable when the repeated measures are indexed by time since adjacent pairs of measures will almost always be more correlated than pairs separated by a greater time interval. If the investigator does not wish to prejudge the validity of the equal correlation assumption, a test of equal correlation is available (Box, 1950). When the assumption of equal correlations clearly does not hold, Box suggested that it might hold if the analysis were performed on differences between adjacent measurements rather than on the original data. The only other approach to salvage the univariate analysis of variance models for the analysis of longitudinal data when the equal correlation hypothesis is not tenable is an approximate procedure proposed by Greenhouse and Geisser (1959). They have shown that the ratios of mean squares have approximate

F-distributions with modified degrees of freedom which are a function of the unknown population variance-covariance matrix. They further show that there is a lower bound on the degrees of freedom which is independent of the unknown parameters. Unfortunately, the use of this lower bound gives a test which is conservative in the sense that the null hypothesis will too often be accepted when it is not true. This loss of power may well be unacceptable.

Because of the restrictive assumptions of equal variances and covariances, it is clear that univariate analysis of variance approaches are not applicable in most situations and that other models which are not dependent on this assumption are needed. Multivariate analysis of variance techniques provide such a class of models. In the case of a single sample of individuals measured longitudinally, the multivariate analog of trend analysis can be thought of as a multivariate generalization of a paired t-test. In this situation, the data consist of the vectors $\underset{\sim}{x}_i' = (x_{i1}, x_{i2}, x_{ip})$ for i=1, ...,N and the hypothesis of interest is that of no time effect, i.e.,

$$H_o: \underset{\sim}{\mu} = \mu \underset{\sim}{j}$$

where $\underset{\sim}{\mu}$ is the mean vector, μ is a scalar, and $\underset{\sim}{j}' = (1, 1, \ldots, 1)$. Morrison (1972) has shown that under the assumption that the observations are an independent sample from a multivariate normal distribution, the maximum likelihood test of this hypothesis is equal to a test of the hypothesis

$$H_o: \underset{\sim}{C}\underset{\sim}{\mu} = \underset{\sim}{0}$$

where $\underset{\sim}{C}$ is any (p-1) by p matrix with the property $\underset{\sim}{C}\underset{\sim}{j} = 0$. In practice, $\underset{\sim}{C}$ is chosen so that the transformed observations are the successive differences of the original data.

This model also allows the investigator to obtain simultaneous confidence intervals for all contrasts $\underset{\sim}{a}'\underset{\sim}{\mu}$ of the repeated treatment means using Scheffé's method of multiple comparisons. When the null hypothesis is rejected, this capability allows more precise statements to be made about the nature of the time response. In particular, the contrasts $\underset{\sim}{a}$ could be chosen to be orthogonal polynomial contrasts. In addition to this basic result, Morrison (1972) derives analogous test statistics and confidence intervals under the more restrictive assumptions of (1) equal variances and covariances and (2) reducible form for the variance-covariance matrix. He then compares the lengths of the confidence intervals with those derived with no structural assumptions. It seems clear that these methods provide a reasonable alternative to trend analysis under a variety of conditions which an investigator might be willing to assume.

The multivariate analysis of variance approach to the k-sample problem of repeated measures is known in the literature as profile analysis (Greenhouse & Geisser, 1959; Morrison, 1967). The basic model is that of a k-sample multivariate analysis of variance in which the observation on the jth individual in the ith group is denoted $\underset{\sim}{y}'_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijp})$ and is assumed to have a multivariate normal distribution with mean $\underset{\sim}{\mu}_i$ and variance covariance matrix Σ. The linear model for these observations is

$$\underset{(n\times p)}{E\,(\underset{\sim}{Y})} = \underset{(n\times k)}{\underset{\sim}{X}}\ \underset{(k\times p)}{\underset{\sim}{B}}$$

where $\underset{\sim}{X}$ is the k-sample design matrix and $\underset{\sim}{B}$ is the matrix of group means. In the context of this model, it is possible to test hypotheses of the form $\underset{\sim}{C}\ \underset{\sim}{B}\ \underset{\sim}{A} = \underset{\sim}{\Gamma}$ for arbitrary $\underset{\sim}{C}$, $\underset{\sim}{A}$, and $\underset{\sim}{\Gamma}$ satisfying the requirements of the general Gauss-Markoff theorem. In particular there exist choices of $\underset{\sim}{C}$, $\underset{\sim}{A}$, and $\underset{\sim}{\Gamma}$ to test the three basic null hypotheses of interest.

The first of these null hypotheses H_{01}, is that the k profiles are parallel which is analogous to the test of no group by time interaction in the univariate analysis of variance approach. For appropriate choice of $\underset{\sim}{C}$, $\underset{\sim}{A}$ and $\underset{\sim}{\Gamma}$, this hypothesis has the form

$$H_{01}: \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{23} \\ \vdots \\ \mu_{1(p-1)} - \mu_{p} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{k1} - \mu_{k2} \\ \mu_{k2} - \mu_{k3} \\ \vdots \\ \mu_{k(p-1)} - \mu_{kp} \end{bmatrix}$$

which is equivalent to a one way multivariate analysis of variance on the differences between measures made at adjacent times.

The second null hypothesis H_{02} is that there is no change through time, that is,

$$H_{02}: \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{k1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \vdots \\ \mu_{k2} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{1p} \\ \vdots \\ \mu_{kp} \end{bmatrix}$$

Matrices $\underset{\sim}{C}$, $\underset{\sim}{A}$ and $\underset{\sim}{\Gamma}$ can be found to test the hypothesis in this form which assumes nothing about the parallelism of the profiles. However, Morrison (1967) proposes an alternative choice of the test matrix which causes the hypothesis tested to be based on equality of sums over groups for each variable. The test in this form is

$$H'_{02}: \sum_{j=1}^{k} \mu_{j1} = \sum_{j=1}^{k} \mu_{j2} = \ldots = \sum_{j=1}^{k} \mu_{jp}$$

which is interpretable only under the assumption of parallel profiles.

The third hypothesis, H_{03}, is that there are no group differences. Without the assumption of parallel profiles, this hypothesis has the form:

$$H_{03}: \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{k1} \\ \vdots \\ \mu_{kp} \end{bmatrix}$$

As above, under the assumption of parallel profiles, Morrison (1967) suggests an alternative hypothesis based on the sums over measurements which has the form

$$H'_{03}: \sum_{j=1}^{p} \mu_{1j} = \sum_{j=1}^{p} \mu_{2j} = \ldots = \sum_{j=1}^{p} \mu_{kj} ,$$

As in the case of univariate analysis of variance, the basic k-sample MANOVA model can be generalized to more complex designs by considering the k-samples as levels of a factorial or other experiment or by assuming some structure for the repeated measures. McCall and Applebaum (1973) present such a generalization with six repeated measures structured as a two by three factorial design. They then compare the univariate and multivariate results for this case and conclude that the multivariate approach is superior.

In the profile analysis model, it is important to point out that the only assumptions made are that the longitudinal series for each individual has a multivariate normal distribution with the same variance covariance matrix in each of the k groups. The assumption of parallel time response functions in the k groups is not necessary. It should additionally be pointed out that the model does not assume anything about the structure of repeated measures. It is in fact not necessary that they be indexed by time of measurement, be equally spaced, or even ordered. Because of this lack of structure the model simply tests whether the time response functions have the same shape without providing a model which describes the shape of the function.

The final major class of models which we will consider are the polynomial growth curve (PGC) models. This class of procedures differs from those already considered because the models are formulated as a function of the structure of the repeated measures. In the previous cases, this structure could be incorporated by considering appropriate contrasts but was not included in the overall tests of hypotheses.

The development of the PGC approaches goes back to the pioneering work of Wishart (1938) which was alluded to earlier as a way of summarizing an individual's time response with a few lower order orthogonal polynomial regression coefficients. Rao (1958) improved on this basic idea by suggesting that the time scale be transformed so that more complex time response functions could be adequately summarized by the linear coefficient computed with respect to the modified time axis. More recent developments in the area of estimating and testing hypotheses about the average PGC of one or more groups have been provided by Rao (1959, 1965, 1966), Potthoff and Roy (1964), Khatri (1966) and Grizzle and Allen (1969). These investigators have provided a variety of procedures which are equivalent under certain but not all conditions. Because of the extent of overlap between approaches, we will concentrate primarily on the Potthoff and Roy approach since their basic model seems most appealing. However, we will point out relationships between their results and the work of Khatri and Rao.

As presented above, the usual MANOVA model can be written as

$$E\underset{(n\times p)}{(\underset{\sim}{Y})} = \underset{(n\times m)}{\underset{\sim}{X}}\ \underset{(m\times p)}{\underset{\sim}{B}}$$

where the rows of $\underset{\sim}{Y}$ are assumed to be independent and follow a multivariate normal distribution with variance covariance matrix Σ, $\underset{\sim}{X}$ is a design matrix of known constants, and $\underset{\sim}{B}$ is a matrix of unknown parameters. In the context of this model, it is possible to test hypotheses of the form

$$H_0:\ \underset{(q\times m)}{\underset{\sim}{C}}\ \underset{(m\times p)}{\underset{\sim}{B}}\ \underset{(p\times u)}{\underset{\sim}{A}} = \underset{(q\times u)}{\underset{\sim}{\Gamma}}$$

for appropriate choices of $\underset{\sim}{C}$, $\underset{\sim}{A}$ and $\underset{\sim}{\Gamma}$ satisfying the generalized Gauss-

Markoff theorem (Timm, 1975). It is also possible to provide simultaneous confidence intervals for functions of the form

$$\underset{(1\times q)}{\underset{\sim}{b}'}\quad \underset{(q\times m)}{\underset{\sim}{C}}\quad \underset{(m\times p)}{\underset{\sim}{B}}\quad \underset{(p\times u)}{\underset{\sim}{A}}\quad \underset{(u\times 1)}{\underset{\sim}{f}}$$

for all $\underset{\sim}{b}$ and $\underset{\sim}{f}$.

Potthoff and Roy (1964) propose a more general model of the form

$$E\ \underset{(n\times q)}{(\underset{\sim}{Y}_o)} = \underset{(n\times m)}{\underset{\sim}{X}}\quad \underset{(m\times p)}{\underset{\sim}{B}}\quad \underset{(p\times q)}{\underset{\sim}{Q}}$$

where $\underset{\sim}{Y}_o$ has rows which are independent and follow a multivariate normal distribution with variance-covariance matrix Σ_o , $\underset{\sim}{X}$ is the 'between individual' design matrix of known constants, $\underset{\sim}{B}$ is a matrix of unknown parameters, and Q is the 'within individual' design matrix. Potthoff and Roy show that this model can be reduced to the previous MANOVA model with the same parameter matrix $\underset{\sim}{B}$ by considering the transformed variable

$$\underset{\sim}{Y} = \underset{\sim}{Y}_o\ \underset{\sim}{G}^{-1}\ \underset{\sim}{Q}'\ (\underset{\sim}{Q}\ \underset{\sim}{G}\ Q')^{-1}$$

where $\underset{\sim}{G}$ is an arbitrary q by q symmetric positive definite matrix. In their original discussion, Potthoff and Roy suggested that the choice $\underset{\sim}{G} = \underset{\sim}{\Sigma}_o$ would be optimum but that since $\underset{\sim}{\Sigma}_o$ was unknown and the distribution theory of using a data derived estimate of $\underset{\sim}{\Sigma}_o$ was unknown, another choice which approximated $\underset{\sim}{\Sigma}_o$ but which was not data based would be appropriate. The choice of taking $\underset{\sim}{G} = \underset{\sim}{I}_q$ was also discussed. Subsequent results by Khatri (1966), Rao (1965), and Lee (1974) established the usefulness of choosing $\underset{\sim}{G} = \underset{\sim}{\hat{\Sigma}}_o$ where $\hat{\Sigma}_o$ is the data based estimate of Σ_o.

Given this basic model, the one sample problem considered previously can be parameterized by choosing

$$\underset{(n\times 1)}{\underset{\sim}{X}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{and } \underset{\sim}{Q} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_q \\ \vdots & & & \\ t_1^{p-1} & t_2^{p-1} & \cdots & t_q^{p-1} \end{bmatrix}$$

so that the expected value of the jth observation on the ith subject has the form

$$E(y_{ij}) = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \dots + \beta_p t_j^{p-1}$$

for all i and j. It is important to point out that the form of the time response is assumed to be the same for each subject i.e., have the same degree. For appropriate choices of matrices $\underset{\sim}{C}$ and $\underset{\sim}{A}$, this model allows an investigator to test hypotheses about the regression coefficients. In particular, one could test the adequacy of a model of a certain degree; or using the result for simultaneous confidence intervals, confidence bounds for the mean growth curve could be derived.

The generalization of this model to the case of k groups of individuals with N_i individuals in the ith group is straightforward. The matrix X(N,k) is constructed to contain N_1 rows of (1,0,...,0), N_2 rows of (0,1,0,...,0), ..., and N_k rows of (0,0,...1). The matrix Q is chosen as above. With this specification, the expected value of the jth observation on the ith subject in the kth group has the form

$$E(y_{kij}) = \beta_{k1} + \beta_{k2} t_j + \beta_{k3} t_j^2 + \dots + \beta_{kp} t_j^{p-1}$$

With this model, matrices C and A could be chosen to test the complete equality of the k regressions, the parallelism of the regressions, or the adequacy of a model of some lower degree. As in the case of the other methodologies discussed, generalizations of the k-sample model to more complex situations are possible.

Timm (1975) presents an example in which the k groups correspond to levels of a two factor factorial experiment.

Alternatives to the use of the Potthoff-Roy model include the one sample PGC model proposed by Rao (1959) and the independent but essentially complementary developments by Khatri (1966) and Rao (1965, 1966, 1967) which are argued to be superior since they eliminate the arbitrary choice of the matrix G. These models have the form

$$E\,(\underset{\sim}{Y}) = \underset{\sim}{X}\;\underset{\sim}{B} + \underset{\sim}{Z}\;\underset{\sim}{\Gamma}$$

where $\underset{\sim}{Y}$, $\underset{\sim}{X}$, and $\underset{\sim}{B}$ are as before, $\underset{\sim}{Z}(N,p-q)$ is a matrix of covariates chosen from the higher order orthogonal polynomial coefficients, and $\underset{\sim}{\Gamma}$ is a matrix of unknown covariate coefficients. If the covariates are not included, the results are identical to the choice G = I in the Potthoff-Roy formulation. If all of the q-p covariates are used,the Rao model is equivalent to the model proposed in Rao (1959) and to the choice G = S in the Potthoff-Roy model where S is the data estimate of Σ_o. Rao (1966) and Grizzle and Allen (1969) recommend the use of some but not all of the p-q possible covariates with the decision of which covariates to include determined by the data. The important point of this rather technical discussion is that the various choices are more similar than different and that each formulation has its problems, i.e., the choice of G for Potthoff and Roy and the choice of which covariates for Rao. In any case the class of models is rich and seems to answer most questions of interest.

The preceeding sections discussed a variety of methods for interindividual analysis which (1) were derived under an assumption that the data were sampled from a univariate or multivariate normal distribution (2) made inferences in classical statistical fashion on the basis of the sampling distribution of statistics (3) made a tacit assumption that all data were present in all of the longitudinal series studied, and (4) considered only a single attribute measured longitudinally. It is the purpose of this section to discuss briefly approaches in the literature for which not all of these four conditions pertain.

The assumption of univariate normality usually does not present a problem in most data analyses since the validity of assuming normality or the extent of deviation from normality can be assessed easily either by using a testing procedure or by inspecting histograms or probability plots. In contrast, the assumption of multivariate normality raises more serious problems since testing and graphic procedures are not nearly as available and results concerning the robustness of procedures in the absence of normality are largely unknown (Kowalski, 1972). In order to avoid the assumption of normality, nonparametric approaches have been developed for many data contexts, but these approaches have been conspicuously absent from the longitudinal data analysis literature. One exception is the paper by Ghosh, Grizzle, and Sen (1973). In this paper, two examples are considered in which the longitudinal series for each individual are replaced by a vector of regression coefficients which summarize the individual's time response function. Under the assumption that these coefficients have a continuous but not necessarily multivariate normal distribution, statistics based on ranks of the coefficients are proposed, and inferences are based on the permutation distributions of these statistics which are asymptotically χ^2. The main hypothesis tested is equality of treatment groups in a design that includes a block factor. While the precise results on the asymptotic relative efficiency of these procedures are not known, the authors assert that these approaches have high asymptotic relative efficiencies for distributions with heavy tails and that the procedures are robust in the presence of gross errors or outliers.

The second common attribute of the interindividual procedures discussed in the previous sections is the fact that they are all based on sampling distributions of statistics. This fact is not in any sense an assumption of the models analogous to the assumption of normality but rather a constraint on the type of inferential statements that can be made. The major alternative inferential context is the Bayesian approach which, among other differences, allows prior information about parameters to be formally incorporated in the analysis. Unfortunately, there are few examples of the application of Bayesian inferential methods to longitudinal data analysis problems. One such application is the work of Geisser (1965) and Geisser and Kappenman (1971). In these papers, the profile analysis model is considered from a Bayesian point of view for, respectively, the two and k-group cases. Under the assumption of parallel profiles, a posterior region is derived for the difference between profiles in the case of two groups and for the vector of differences between the k-1 pairs of adjacent profiles in the case of k-groups. This derivation is considered for both "non-informative" and "natural conjugate" priors.

The third point concerns the presumption of complete longitudinal series. This requirement is, in practice, quite severe since it is often the case in protracted studies that only a small percentage of the series are complete for all ages. The loss of data imposed by this constraint is even more serious inferentially if there is any reason to believe that the occurrence of missing data is in any way related to the value of the attribute being measured. This problem can be dealt with in at least a couple of ways. One method would be to take the approach used by Wishart (1938) to replace the longitudinal series by summary parameters which can be estimated even in the presence of a moderate amount of missing data. Such derived data, though not precisely identically distributed, should allow at least an approximate analysis using a larger sample size.

A more formal approach to this problem was suggested in a recent paper by Kleinbaum (1973) who generalized the polynomial growth curve formulation of Potthoff and Roy to consider the presence of missing data. In the presence of complete data the model has the form

$$\underset{(N\times q)}{E\,(\underset{\sim}{Y})} = \underset{(N\times m)}{\underset{\sim}{X}} \quad \underset{(m\times p)}{\underset{\sim}{B}} \quad \underset{(p\times q)}{\underset{\sim}{Q}}$$

If the structure of the data is such that there are ℓ blocks of cases with N_ℓ cases in block ℓ and that within block ℓ all cases are complete for some number q_ℓ of the q observations, Kleinbaum proposes a modified model

$$\underset{(N_\ell\times q_\ell)}{E\,(\underset{\sim}{Y}_\ell)} = \underset{(N_\ell\times m)}{\underset{\sim}{X}_\ell} \quad \underset{(m\times p)}{\underset{\sim}{B}} \quad \underset{(p\times q)}{\underset{\sim}{Q}} \quad \underset{(q\times q_\ell)}{\underset{\sim}{H}_\ell}$$

where $\underset{\sim}{H}$ is an incidence matrix of zeros and ones. With this model it is possible to obtain best asympototically normal estimates for linear functions of the parameters and to test hypotheses about such linear functions.

While this approach may be useful in correcting for data missing by chance, it is also applicable to situations in which data are missing by design as in the case of mixed longitudinal cross-sectional designs (Prahl-Andersen & Kowalski, 1973).

The fourth point concerns the fact that all of the preceeding discussion at both the intra- and inter-individual levels has been restricted to situations which are univariate in the sense that the data have consisted of a series of measurements of a single attribute indexed by time. The extension of these approaches to the case of a three dimensional data matrix in which two or more variables are measured longitudinally introduces a new level of complexity.

Several approaches to this problem have been suggested in the literature for both intra- and inter-individual analyses (Kowalski & Guire, 1974). Of these, the approach most widely used in biological applications is bivariate allometry which relates the growth of exactly two dimensions in a single sample of cases. Attempts to extend this approach to more than two dimensions have been made but not without introducing additional problems of interpretation. Another avenue of approach to this problem has been in the area of factor analysis generalized to the case of a three dimensional data matrix. Such approaches, which go beyond the scope of this presentation, seem also to introduce difficult problems of interpretation.

Of the topics discussed in this paper for the univariate case, two areas seem to offer a way of approaching the problem of a three dimensional data matrix. The first approach is simply to reduce the problem to a two dimensional one by summarizing the longitudinal series for each variable with one or more derived variables. The methods of section III for intra-individual analysis provide a variety of possible ways in which this could be done. Possible candidates for such summary variables include orthogonal polynomial coefficients (Wishart, 1938; Rao, 1958), the parameters of an appropriate Gompertz or logistic model (Bock et al., 1973), or the scores derived from a principal components analysis of the longitudinal series as suggested by Rao (1958). Such summary parameters could then be used in a variety of multivariate analyses which either analyze the structure of a single sample or compare two or more samples. The utility of this approach obviously depends on the choice of summary variables, which introduces a certain degree of subjectivity into the analysis. However, it would seem that this approach makes considerable data analytic sense.

The second approach is provided by the Potthoff and Roy polynomial growth curve models which can be applied directly to the case of two or more variables measured concurrently. This can be done simply by appropriate choice of the pre- and post-design matrices. One could, for example, specify a model in which a polynomial was fit separately for two or more variables taking into account not only the correlations within a series but the correlations between series as well. Having fit such a model, one could test whether the several time response functions were equal or parallel. More complex models involved more than a single sample could also be considered.

SUMMARY

We have attempted to survey a variety of methods which are appropriate for the analysis of a single longitudinal series and for the analysis of one or more samples of longitudinal observations. We also attempted to place in perspective the role of such methodologies in the broader context of developmental research. Having done this, it seems appropriate to comment on the current state of the art from the point of view of both theory and practice.

In 1963 Bereiter, observed that deficiencies of statistical methodology seriously impaired the scientists investigation of questions dealing with change. Since that observation was made, a great deal of theoretical work has been carried out. At the intra-individual level, new models have been proposed by Bock (1973) and others which are parameterized in ways that facilitate biological interpretation of the fitted curve. At the inter-individual level, the development of polynomial growth curve models which began with the work of Rao (1959) and Potthoff and Roy (1964) is certainly the most notable advance of the last few years. Because of these achievements and others one would have to conclude that the state of the art has indeed improved in recent years. One would also have to conclude that there are many interesting and challenging problems remaining. At the theoretical level, the problems of growth prediction for individual series, of multivariate data observed longitudinally, and of nonparametric alternatives to normal theory procedures stand out as areas of ongoing interest. At the applied level, the challenge of testing new methodologies in a variety of contexts always exists as statistical practice lags frustratingly far behind statistical theory.

ASSESSMENT AND STATISTICAL CONTROL OF SUBJECT VARIABLES IN LONGITUDINAL DESIGNS

Paul A. Games
The Pennsylvania State University

Table of Contents

ASSESSMENT AND STATISTICAL CONTROL OF SUBJECT VARIABLES IN LONGITUDINAL DESIGNS

ABSTRACT

I. INTRODUCTION

II. REPRESENTATION OF EXPERIMENTAL DESIGNS

III. POSSIBLE ROLES OF ORGANISMIC VARIABLES

IV. COVARIANCE TYPE MODELS WITH A NESTED FACTOR AND A REPEATED MEASURE, BUT WITH X VALUES AVAILABLE ONLY ON SUBJECTS, NOT ON EACH OBSERVATION

V. COVARIANCE TYPE MODELS WITH ONE COVARIATE VALUE PER OBSERVATION (SEVERAL PER SUBJECT)

VI. CONCLUSIONS AND SUMMARY

ABSTRACT

An experimental design notational system is described. A "minimal power" check is recommended for use with complex analyses of covariance structures. Longitudinal models always include repeated measures. Conventional ANOVA longitudinal analyses are contrasted with covariance type linear models. The covariance models have the virtue of permitting the use of P different organismic (X) variables, while conventional ANOVA usually is limited to one. Tests of homogeneity of regression slopes are illustrated for covariance type models with a repeated measure factor and: (a) X's available on subjects but not each separate occasion; or (b) X's available for each occasion.

I. INTRODUCTION

Longitudinal studies automatically imply that there will be one or more repeated measure factors (Kirk, 1968; Winer, 1971). For convenience we shall label one repeated measure factor as T, the time factor. The simplest possible design is to obtain several subjects drawn at random from some meaningful population, and record the dependent variable, Y, upon those subjects over the T factor. This permits the plotting of individual "growth" or T curves on Y. Unfortunately, when we note that the curves for John Smith and for Tom Johnson are different, we have gained little useful knowledge. Since these individuals differ in many ways, we have no basis for distinguishing future individuals who are likely to show the "Smith curve" instead of the "Johnson curve". To obtain useful information it is necessary to have available data on other trait or organismic variables that may be used to classify or group the individual subjects. We shall label these variables as X's. X may be something as easily observed as sex, or something that must be measured by instrumentation or psychological tests. Thus we may have a study in which children are grouped into those with internal locus of control versus those with external locus of control. If these groups now show divergent T curves on Y, we have obtained useful information such that other children may be assessed on this X and predictions made about the type of T curve expected. Thus the minimum useful design is one where there is at least one X in addition to the dependent variable Y.

II. REPRESENTATION OF EXPERIMENTAL DESIGNS

It is helpful to have a concise notation to express the information available in a given design. We need two basic terms. Two factors are crossed if each level of each factor appears with each level of the other factor. Since each subject, S, appears with each level of T, we say S and T are crossed and represent this as S x T. We may use subscripts on the factors to represent the number of levels. Thus S_n x T_5 indicates that n subjects are crossed with 5 levels of time.

If a level of factor A occurs within only one level of factor B, then we say A is nested in B. This is represented as A in (B). In a completely randomized one factor design where 50 subjects are randomly divided into 10 subjects under each of 5 levels of factor A, we would write S_{10} in (A_5). A completely randomized factoral design with 2 levels of A, and 3 of B may be represented as S_n in (A_2 x B_3). Since subjects are nested in the AB crossing, this shows that any one subject appears in only one of the 6 cells formed by the AB crossing. Lee (1975) has written an ANOVA text using crossing and nesting notations throughout that should be consulted if an expansion of this brief explanation is desired.

We must also add a notation to express the role of X. We shall use X to stand for any continuous variable available other than the dependent variable, Y. We shall arrange the terms so that when X is present, it implies an X in every unit prior to the appearance of X. Thus S_n w X x T_4 implies that there is an X available with every subject, or here a total of n values of X. However, if there is an X available on each observation, this would be represented by placing the "w X" last. Thus S_6 x T_4 w X implies there are 24 values of X available, or an X associated with each of the 24 observations of Y. The w is for mnemonic value since the S w X x T would be read as Subjects, with an X value on each, crossed with T.

These symbols may be combined to represent more complex designs. As an example with one repeated measure, B, and one between subjects measure, A, Winer has an [S_3 in (A_2)] x B_4 design (1971, p. 525). The [ ] brackets indicate B is crossed with both A and subjects. We must distinguish whether there is only one X for each subject, or one X for each observation. An example of the former is an $[S_4 \text{ w } X \text{ in } (A_2)] \times B_2$ in Winer (1971, p. 803) while an example of the latter is an $[S_3 \text{ in } (A_3)] \times B_2 \text{ w } X$ (Winer, 1971, p. 806).

Naturally it is possible to have more than one organismic variable in a study. To indicate more than one, we will add extra values of X. We shall designate the number of X variables as P, and may represent a more general case of Winer's covariance example (1971, p. 803) as $[S \text{ w } X_1 \ldots X_p \text{ in } (A_2)] \times B_2$.

III. POSSIBLE ROLES OF ORGANISMIC VARIABLES

A. Reduction of Error Term

There are several possible roles of organismic variables in longitudinal designs. In some designs they may be included primarily as a device to reduce the error term, and increase the power and precision of the comparisons made. If we have variable X, we may divide it into 4 adjacent intervals, and treat these four levels as a factor with subjects nested in the four levels. If the initial design involves a manipulative factor A, the design may be expanded by adding the additional O factor with four levels. Thus an initial $[S_{40} \text{ in } (A_3)] \times T_2$ design may be expanded to an $[S_{10} \text{ in } (A_3 \times O_4)] \times T_2$ design. In the latter design the addition of the O factor should reduce the error term for the A effect, $MS_{S(AO)}$, just as in a treatment by levels design. The statistical considerations of such usage are well covered in Lindquist (1953, chapter 5) and Myers (1972, chapter 6).

An alternative way of using an organismic variable to reduce mean square error (MS_E) is to use X as a covariate in a legitimate experimental design. With an $[S \text{ w } X \text{ in } (A)] \times T$ design where subjects were randomly assigned to the A conditions so the $\bar{X}$ values vary only by chance, the major role of a covariate would be in reducing $MS_{S(A)}$ term, the error term for the A main effect. A test for heterogeneous slopes in such a design is illustrated later in the present chapter.

B. Identifying Variable by Time Interactions

A second role of organismic variables is in a search for organismic variable by time interactions, or a search for organismic variable by treatment interactions that occur over time. In an [S in (A x O)] x T design, the AO interaction would be an example of an "aptitude by treatment" interaction. The presence of an OT interaction would indicate different growth curves for the several O levels. A significant AOT interaction would indicate that the aptitude by treatment interaction changed over time.

Similar searches for O interactions are possible without subdividing X into levels and using it as an additional factor in an analysis of variance. In an [S w X in (A)] x T design the check of homogeneous regression slopes is equivalent to a test for a linear AO interaction. Thus the finding of heterogeneous slopes in a covariance model is another way of discovering an "aptitude by treatment" interaction.

Similarly, if there is interest in a possible X by T interaction, it is possible to build vectors consisting of the X vector multiplied by contrasts of the T factor to determine if there is an XT interaction. Unfortunately, this involves methodology that is less likely to be used. Similarly, the search for a triple interaction by use of AXT vectors is even less common. The beginning student is probably well advised to use the blocking method of forming an O variable, and AOV when he has only a single X variable. The complexity of the other methods is mainly needed when there are several X variables.

C. Clarifying the Nature of Relationships

The third role of organismic variables is in an attempt to "eliminate" the effect of other "extraneous" organismic variables to provide "clear" interpretations. This is the role of partial correlations, part correlations, and sometimes of analysis of covariance when the $\bar{X}$'s differ systematically due to the fact that a legitimate randomization of experimental subjects has not been carried out. Unfortunately, such usage is easy to misinterpret. Campbell and Erlebacher (1970) give a long, detailed presentation of such problems.

Some such studies take a multiple regression form. We may ask, for example, how much effect does education have on income at age 40 after IQ, SES, and parental income have been "partialed out." Such a study computes the residual increment in $R^2_{Y.1234} - R^2_{Y.123}$ where education is the fourth variable of the four highly intercorrelated variables. The problem is, each of the four variables may have a very low "additional contribution" so that no matter which of the variables is placed last, it will contribute very little once the other variables have entered. Thus $R^2_{Y.1234} - R^2_{Y.124}$ may be equally low suggesting that parental income is a very minor contribution once education, SES, and IQ have been partialed out. Unfortunately, some authors (e.g., Bowes & Gintus, 1972) report only the one type of partialling that fits the investigators' intellectual framework. The statistically naive reader is left with the impression that this has been a dramatic demonstration that education has no effect on future income. Such a study merely reveals high intercorrelations between the predictors, so that no one predictor makes much of a unique contribution once the others have been used.

Unfortunately, there is no solution to obtaining clear causative conclusions in the absence of manipulative studies. Studies in which subjects can not be assigned at random can yield tentative guesses of causative chains, but no more. Sometimes, these tentative guesses can be strengthened by attempts that show the same trends exist even after a possible extraneous uncontrolled factor has been "partialed out." Thus if an [RS in (O)] x T experiment shows that low anxiety subjects have a more rapidly rising learning curve than high anxiety groups, the interpretative or theoretical value of this result may be questioned when it turns out that the low anxiety group had a mean IQ of 110 and the high anxiety group had a mean IQ of 100. "Partialling out" the IQ variable may restore part of the credibility of the finding, if the same results are obtained. However, such results never have the interpretative clarity that can be obtained when "anxiety" is determined from a manipulated condition rather than a biographical inventory. If several manipulative conditions all supposed to produce anxiety yield consistent results, clear causative conclusions are possible. Unfortunately, longitudinal studies must often use organismic variables to designate their groups. From such studies many alternative explanations are always possible.

D. Use of Organismic Variables in Complex Structures

A fourth use of organismic variables here must be a catch all category. We shall call it use of organismic variables in complex structures. This is meant to include factor analysis, analysis of covariance structures, and the models Jöreskog identifies as LISREL models (Jöreskog & Sörbom, 1976a; see also chapter this volume). The latter refers to a very general model that Jöreskog has incorporated into a computer program that permits specification and testing of: a measurement model of latent variables in the X's, a measurement model of latent variables in the Y's, and a structural analysis relating the X latent variables and the Y latent variables. This general structure includes many multivariate analyses as special cases. It permits specification of simplex models on the Y's with X's as additional predictors.

The numerous possibilities are too many, and too complicated to comment on here, except for a brief philosophical note. We will be seeing many complex models formulated on behavioral data sets. In many cases, the authors report a given model, say it is compatible with the data, and let it go at that. Jöreskog's programs permit maximum likelihood tests of specified models. It is hoped that readers will learn to take such tests in a sensible fashion, without some of the habits that seem to exist in the interpretation of simpler hypotheses tests. One of the most superficial review practices is to take a given area of content, or a given experimental question, then to obtain all references on that topic and tabulate the number of significant differences. In asking "Is A greater that B," some review articles say 15 were significant with A>B, 25 were nonsignificant, so A = B was retained, and 2 were significant with B>A. Therefore the A = B's have it, and further research is needed. Such reviews often ignore the fact that many of the studies may have had such small n's that for any reasonable A-B effect, the power was minimal, and retention of the null is quite likely. Until reviewers consider power problems, and the adequacy of the experimental design such compilations are of little value. A modicum of statistical sophistication is needed to yield worthwhile reviews that separate the wheat from the chaff, the signal from the noise.

As we go to more complex models, for more complex questions, still more sophistication will be needed. It is clear that for most sets of data, many different models could be built. If n is small enough, the power will be small so that almost any model will "fit" in that the null will be retained. On the other hand, if n is large enough, almost any model will not fit, in that the chi square on the model fit will be significant because some specified parameter in the model will be a little bit different in reality than it is in the model. Readers who use the "significant versus nonsignificant" gauge as their sole evaluation tool for published research are going to be hopelessly lost when it comes to the use of complex models. It is difficult to provide guidance rules that will always apply, but the author would like to add one suggestion.

It is often very difficult to express the adequacy of fit of these complex models. The more interesting publications are often those that at least explore alternative models. Madaus, Woods, and Nuttall (1973) present an interesting study based on a very large n, on the hierarchial model of Bloom's taxonomy of cognitive objectives. They construct a model that knowledge items must be mastered before you can master comprehension items, and one must master comprehensive items before he can master application items on that topic, etc. They compare a model without a general intelligence factor, and a model with a general IQ factor, and conclude that the latter is needed to fit the data. Basically, their final model supports the notion of a hierarchial structure from knowledge to comprehension to application. However, no heirarchy is indicated for analysis, synthesis, or evaluation. Their study is especially interesting because it compares several feasible models with each other. (Further statistics and further models to be included may be desired by the reader; this is a very tough problem for the editors and reviewers of studies testing complex structures. We can only ask that authors make copies of their data available to those who request it.)

To provide at least a minimum basis for publication of such studies, it would be desirable to know that the experiment at least had enough power so that it could reject some "outrageous models." In the Madaus et al. (1973) study, it would be possible to reverse the hierarchy and test a model in which evaluation, synthesis, and analysis are taken as prerequisites to successful mastery of knowledge and comprehension items. That is, an "outrageous" model might be one in which the direction of the hierarchy is reversed. If the study had so little power it was unable to reject this "outrageous" alternative, there certainly is little basis for taking the model that was retained very seriously. As we get complex models, we must recognize that many, many different models may adequately fit the same data. We can only hope to improve our models, much as the physical sciences have done. We must not consider that every model that is not rejected is true, or that every model that is rejected is useless.

Rather than continue in this general vein, let's turn to the class of models probably used most often in longitudinal studies, covariance type models. The author is somewhat skeptical about interpretations of such models when the

$\bar{X}$'s differ considerably (Games, 1976; Campbell & Erlebacher, 1973), but feels that a major problem with covariance usage is that many, or even most, users readily adopt a covariance model without first looking at the more general models that permit heterogeneous slopes. Even modern computer programs (e.g., BMDP2V) do not provide a test for the homogeneous slopes assumption in covariance. The following two sections shall illustrate such tests in repeated measure type designs.

IV. COVARIANCE TYPE MODELS WITH A NESTED FACTOR AND A REPEATED MEASURE, BUT WITH X VALUES AVAILABLE ONLY ON SUBJECTS, NOT ON EACH SEPARATE OBSERVATION

In this design, we have only one set of X's per subject. It is possible to have several variables in X, as in $[RS_n w\ X_1 \ldots X_p \text{ in } (A_a)] \times B_b$ and such designs would be carried out using the steps we will illustrate below, only with several vectors of X. For convenience, we shall use data with a single X vector. Winer (1971, p. 803) has such a set of data analyzed by covariance. One of the assumptions of covariance is that the regression coefficients within groups is the same. Winer does not illustrate how to test this assumption on any designs with repeated measures nor does any other source the author knows of. To illustrate the procedures needed in a more general context, we shall add a third (A_3) group of independent subjects to Winer's data, thus resulting in an $[RS_4 w\ X \text{ in } (A_3)] \times B_2$ design. B is here used, instead of T, to match Winer's symbolism.

A. Linear Models for a Test of Heterogeneous Regression Slopes

The data are given in Table 1, with various vectors needed for a linear models solution with homogeneous or heterogeneous slope solutions. The first three columns are the usual subscripts for the design factors. Column four is the dependent variable, Y. There are two Y's for each subject (due to the repeated measure factor B with two levels), here each observation is found in a different row, with the corresponding B level indicated in the third column. The unity vector of all 1's (for estimation of μ_y) is column five. The control variable, X, is given in the sixth column. Since there is only one X per subject, we repeat the X value for each of the two rows of that subject. Thus subject 1 has a 3 punched in both of the first two rows ($X_{11} = 3$). The X value is given twice for all subjects. Since X varies only over the different subjects, it can influence only the results on the between subject factors, A and S(A). Columns seven and eight are vectors expressing the three levels of A in two orthogonal contracts: -2, +1, +1, and 0, -1, +1.

Insert Table 1 about here

Similarly column nine expresses the B +1, -1 contrast as a vector. If there were more than two levels of b, we would need (b-1) such vectors, each an orthogonal contrast. In addition, we need vectors for the subjects, but since these are both numerous and constant for any of the analyses we shall consider, we shall not clutter up the table with them; they are implied (see Cohen & Cohen, 1975, chapter 10). The final vectors needed are for the interaction. When we multiply a main effect contrast vector for A by a main effect contrast vector for B in a balanced design, we obtain an orthogonal interaction vector, as in columns ten or eleven. Thus we have a set of five mutually orthogonal experimental design vectors.

If we use columns 5, 7-11, and the subject vectors as predictors in the X matrix of a multiple regression on Y, we may obtain the usual ANOVA as given in Table 2.

Since the five experimental vectors are mutually orthogonal, we could obtain a SS with one df for each vector and these would sum to SS_{cells}. However, it is conventional to sum the single df into the usual omnibus SS for a factor

Table 1

Data & Vectors for an [RS_4 w X in (A_3)] x B_2 Design (Winer, 1971 p. 802)

Columns

1	2	3	4	5	6	7	8	9	10	11	12	13
A	S	B	Y	U	X	ψ_{A1}	ψ_{A2}	ψ_B	ψ_{A1B}	ψ_{A2B}	$X\psi_{A1}$	$X\psi_{A2}$
1	1	1	10.0	1	3.00	-2	0	+1	-2	0	-6.00	0.0
1	1	2	8.00	1	3.00	-2	0	-1	+2	0	-6.00	0.0
1	2	1	15.0	1	5.00	-2	0	+1	-2	0	-10.0	0.0
1	2	2	12.0	1	5.00	-2	0	-1	+2	0	-10.0	0.0
1	3	1	20.0	1	8.00	-2	0	+1	-2	0	-16.0	0.0
1	3	2	14.0	1	8.00	-2	0	-1	+2	0	-16.0	0.0
1	4	1	12.0	1	2.00	-2	0	+1	-2	0	-4.00	0.0
1	4	2	6.00	1	2.00	-2	0	-1	+2	0	-4.00	0.0
2	1	1	15.0	1	1.00	1	-1	+1	+1	-1	1.00	-1.00
2	1	2	10.0	1	1.00	1	-1	-1	-1	+1	1.00	-1.00
2	2	1	25.0	1	8.00	1	-1	+1	+1	-1	8.00	8.00
2	2	2	20.0	1	8.00	1	-1	-1	-1	+1	8.00	-8.00
2	3	1	20.0	1	10.0	1	-1	+1	+1	-1	10.0	-10.0

Table 1 (cont.)

2	3	2	15.0	1	10.0	1	-1	-1	-1	+1	10.0	-10.0
2	4	1	15.0	1	2.00	1	-1	+1	+1	-1	2.00	-2.00
2	4	2	10.0	1	2.00	1	-1	-1	-1	+1	2.00	-2.00
3	1	1	11.0	1	2.00	1	+1	+1	+1	+1	2.00	2.00
3	1	2	8.00	1	2.00	1	+1	-1	-1	-1	2.00	2.00
3	2	1	15.0	1	6.00	1	+1	+1	+1	+1	6.00	6.00
3	2	2	13.0	1	6.00	1	+1	-1	-1	-1	6.00	6.00
3	3	1	19.0	1	9.00	1	+1	+1	+1	+1	9.00	9.00
3	3	2	15.0	1	9.00	1	+1	-1	-1	-1	9.00	9.00
3	4	1	12.0	1	3.00	1	+1	+1	+1	+1	3.00	3.00
3	4	2	7.00	1	3.00	1	+1	-1	-1	-1	3.00	3.00
Σ			327.		118.	0	0	0	0	0		
Means			13.625		4.917							
Cov b's				7.4167		.45833		-2.125			.375	
					1.250		-2.125			0.0		

or interaction, as given in Table 2. (When apriori reasons specify a particular interaction or main effect vector as of prime interest, it is advisable to test that one vector alone.) The summary table of the ANOVA has ignored the presence of the X vector as a predictor.

Insert Table 2 about here

To obtain an analysis of covariance, we merely include the X vector (column G) in the predictor variable matrix of multiple regression. However, if we insert this _one_ X vector alone, we are assuming that X acts the same way in all three of the independent groups of the A factor. We are assuming that the subject means (summed over the B factor) have the same regression from X in all three groups. Before making such an assumption, it is wise to test it.

To test for homogeneity of regression, we add not only column 6, but also columns 12 and 13 to the predictor matrix. Given that vector 6 is already in the analysis, column 12 is testing H_0: $\beta_1 = (\beta_2 + \beta_3)/2$ where the β's are the raw score regression coefficients in the population for the three respective A populations, A_1, A_2, and A_3. Similarly column 13 is testing H_0: $\beta_2 = \beta_3$. If both of these are true, then $\beta_1 = \beta_2 = \beta_3$; i.e., we have homogeneity of all three regression slopes. The common procedure is to combine these into a single familywise test. This may be carried out several ways, depending upon the computer programs available. If only general multiple regression programs are available, it may be necessary to punch the data as in Table 1, and obtain two different SS (regression) values.

We can simplify the analysis by ignoring the within subject variables. These are orthogonal to all of the between subject variables; and only the latter are influenced by the between subjects covariate vectors in the present design. If we use columns 5, 6, 7, 8, 12, and 13 as predictors for the heter-

Table 2

AOV of Data of

$[RS_4 \text{ w } X \text{ in } (A_3)] \times B_2$ Example

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Squares	F	P
BTN Subjects					
A	2	83.250	41.625	1.213	>.25
S(A) = E	9	308.88	34.319		
Within Subjects					
B	1	108.37	108.37	109.896	<.01
AB	2	2.2500	1.1250	1.141	>.25
SB(A) = EW	9	8.8750	0.98611		
Total (ADJ)	23	511.63			

Y Means, $\bar{Y}_{.jk}$	B_1	B_2	$\bar{Y}_{.j.}$	$\Delta_j = \bar{Y}_{.j1} - \bar{Y}_{.j2}$
A_1	14.25	10.0	12.125	4.25
A_2	18.75	13.75	16.25	5.00
A_3	14.25	10.75	12.50	3.50
$\bar{Y}_{..k}$	15.75	11.50	13.625	

ogeneous A slopes model, we obtain SS (regression) = 339.16 with df = 5. Then dropping to the homogeneous A slopes model, we would use columns 5, 6, 7, and 8 only and obtain SS (Regression) = 328.46 with df = 3. Taking the difference between these two SS we obtain the SS associated with heterogeneous slopes = 339.16 - 328.46 = 10.70. The df is 5 - 3 = 2, so MS (hetero slopes) = 5.35. This should be tested using the adjusted $MS_{S(A)}$ of the heterogeneous slopes model of 8.8269 as MS_E, yielding an F of .606. With an F less than one, we would retain the hypothesis of homogeneous slopes thus justifying adoption of the covariance model.

If the regression slopes were heterogeneous, it would be necessary to proceed using different regression slopes for each of the 3 levels of A. Probably the simplest way would be to obtain the subject means of $\bar{Y}_{ij.}$ (averaged over B) and solve for a separate regression equation for each of the three groups. The procedures are illustrated in Cohen and Cohen (1975, p. 314-319). If you use the simple $\hat{Y} = b_o + b_1 X$ for each group separately, you may insert the value of the grand mean of the X's, $\bar{X}..$, to obtain a predicted value of Y for each of the three groups when X is at its mean value. This plus graphs of the three regression equations will provide useful information. The statistically sophisticated may wish to use the Johnson-Neyman technique as illustrated in Walker and Lev (1953, pp. 398-404), or the extensions discussed by Cahen and Linn (1971).

B. A Linear Model for Covariance

With the above results, it is possible to proceed by covariance, using vectors 5-11. In this case, we obtain the analysis given in Table 3. Note that only the between-subjects factors SS's have changed. Anytime we have only one set of X's per subject, only the between-subject factors may change. In this case, the covariate is successful in reducing the error term, $MS_{S(A)}$ from 34.319 in the AOV to 7.958 in the COV. Thus there will be more power in the

COV than in the AOV. This is the major virtue of COV in a legitimate experimental design. When S's are assigned randomly and X is obtained prior to the treatments, the $\bar{X}_{.j}$'s will differ only by random fluctuation. Hence the adjusted $\bar{Y}_{.j.}$'s for the A groups will not differ from the original $\bar{Y}_{.j}$'s by very much. When the groups are not assigned at random, and the $\bar{X}_{.j}$'s are substantially different, then we get into complexities of interpretation. We are then working on estimates of the effects of A when substantial differences in X have been "partialled out." Interpretations are now similar to those of partial correlations.

Insert Table 3 about here

Note that the adjusted means of the covariance have changed just slightly from the original $\bar{Y}$ means of Table 3. The B means and SS have not changed at all, since the covariance here may change only the between subject effects. The AB interaction SS also is exactly the same, and correspondingly the difference between the two cell means for a given row of A is exactly the same in Table 4 as in Table 3. Only the A main effects have been changed by the covariance (the cell means reflect this effect also). The A main means have changed only a little, because the $\bar{X}_{.j.}$ differ only by small amounts, as would be expected if only random sampling produced the differences. The covariance has now increased the precision so the main A effect is significant whereas it was not in Table 2. The Tukey WSD value for the A means is 1.9697 so the A_2 mean is found to be significantly larger than either the A_1 or A_3 means, the latter two not being significantly different.

There are two different procedures that may be used to solve for the "adjusted A means." Some programs and texts solve for the "adjusted mean" as the predicted Y value using the observed $\bar{X}_{.j.}$ of that particular group. The

Table 3

COV of Data of

$[RS_4$ w X in $(A_3)] \times B_2$ Example

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Squares	F	P
BTN Subjects					
Reg. X_1	1	266.73			
A	2	61.728	30.864	3.879	<.05
S(A)	8	63.663	7.9579		
Within Subjects					
B	1	108.38	108.37	109.896	<.01
AB	2	2.2500	1.1250	1.141	>.25
SB(A)	9	8.8750	0.9861		
Total (ADJ)	23	511.63			

ADJ. Means, $\bar{Y}'_{.jk}$

	B_1	B_2	$\bar{Y}'_{.j.}$	Δ_j
A_1	14.771	10.521	12.646	4.25
A_2	18.333	13.333	15.833	5.00
A_3	14.146	10.646	12.396	3.50
$\bar{Y}'_{..k}$	15.750	11.500	13.625	

present solution uses the grand mean of the X's (4.9167) for all three A groups, so the "adjusted means" are the predicted values of Y for that common point.

C. Linear Models Using More Than One X Vector

One of the virtues of the general linear model approach is that it readily facilitates the use of as many covariates as available, while the conventional procedures covered in Winer (1971), Kirk (1968), Dayton (1970), or Myers (1972) become very awkward with more than one covariant. We shall illustrate the expanded case by adding a second covariate, X_2, to the data of Table 1. The new vector, and the vectors generated from it are found in Table 4. These should be considered a continuation of Table 1.

Insert Table 4 about here

The heterogeneous slopes model would now use all vectors from 5 to 16 as predictors. The homogeneous slopes model includes vectors 5 to 11, plus the new vector 14 ($\underset{\sim}{X}_2$). There are now four vectors included in the heterogeneous slopes model that are not in the homogeneous slopes model. Thus subtracting the SS regression from the two models yields the SS (heterogeneous slopes) = 8.4974 with df = 4. This value, converts to MS of 2.1244. The new value of the adjusted $MS_{S(A)}$ error terms (8.9137) when divided into the above MS yields an F less than 1, so we retain the hypothesis of homogeneous slopes, and would proceed with the usual covariance.

The covariance table, and adjusted means are contained in Table 5. We see that the second covariate has further reduced the MS error term, since it accounts for much of the subject variance (within A). The error term has been reduced from 7.96 in the COV with X_1 alone, to 5.03 in the analysis with both X_1 and X_2 as covariates. However, while the A effect was significant in Table 4, it no longer is significant in Table 5. Partialling out X_2 removes more

Table 4

Additional Vectors Needed With A Second Covariate, X_2

Columns	14	15	16
	X_2	$\psi_{A_1}X_2$	$\psi_{A_2}X_2$
	2.00	-4.00	0.0
	2.00	-4.00	0.0
	7.00	-14.0	0.0
	7.00	-14.0	0.0
	9.00	-18.0	0.0
	9.00	-18.0	0.0
	4.00	-8.00	0.0
	4.00	-8.00	0.0
	7.00	7.00	-7.00
	7.00	7.00	-7.00
	9.00	9.00	-9.00
	9.00	9.00	-9.00
	8.00	8.00	-8.00
	8.00	8.00	-8.00
	5.00	5.00	-5.00
	5.00	5.00	-5.00

Table 4 (cont)

3.00	3.00	3.00
3.00	3.00	3.00
6.00	6.00	6.00
6.00	6.00	6.00
8.00	8.00	8.00
8.00	8.00	8.00
4.00	4.00	4.00
4.00	4.00	4.00

variance from S_{SA} than it does from $SS_{S(A)}$ so the test of the A effect is no longer significant. Those who believe that covariance always will decrease p values are in for disappointments.

Insert Table 5 about here

The observant reader will have noted that the within-subjects effects SS_B and SS_{AB}, have been completely unaffected by all of the above. This is because the A, X_1, and X_2 vectors are all orthogonal to the within-subject effects. Thus analysis of covariance, when there is just a single X set of values for each subject will only influence the between-subjects effects. The within-subjects effects are here reflected by the B main means, (here always 15.75 and 11.5) and the differences between the cell means for each A row (here always 4.25, 5.00, and 3.50). These terms stay the same regardless of any between subject covariate effect. This same difference in between-subjects and within-subjects effects would hold on more complex repeated measure designs exactly as it holds here.

It would be possible to add a third covariate vector to the present set. However, to test for homogeneity of regression would require a total of three additional vectors, as in Table 5, and this would leave zero df for the adjusted error term, $MS_{S(A)}$. In reality, we should have many more subjects to achieve stability when using many covariates.

V. COVARIANCE TYPE MODELS WITH ONE COVARIATE VALUE PER OBSERVATION (SEVERAL PER SUBJECT)

In this situation, we have the entire design available in both the X and the Y values. There is one X paired with each Y. This may be represented in our notational form by placing the w X term after the last term in the design.

Table 5

COV with X_1 and X_2

As Covariates

Source	DF	Sum of Squares	Mean Squares	F
Btn Subjects				
Reg. X_1 & X_2	2	339.373		
A	2	17.509	8.7547	1.739
S(A)	7	35.238	5.0341	
Within Subjects				
B	1	108.38	108.37	96.229
AB	2	2.2500	1.1250	
SB(A)	9	8.8750	0.98611	
Total (ADJ)	23	511.63		

Adjusted Means, $\bar{Y}'_{.jk}$

	B_1	B_2	$\bar{Y}'_{.j.}$	Δ_j
A_1	13.938	9.688	11.813	4.25
A_2	22.187	17.187	19.687	5.00
A_3	11.125	7.625	9.375	3.50
$\bar{Y}'_{..k}$	15.75	11.5	13.625	

This is in contrast to the previous case where w X was placed after the subject term to indicate X's are available only for each subject. Winer (1971, p. 806) has an example design that is an $[S_3 \text{ in } (A_3)] \times B_2$ w X. We shall use these data to illustrate the linear model analyses needed to test for homogeneity of slopes, and the covariance analysis.

Winer's data are given in Table 6. The first three columns contain the subscripts for the A, S, and B factors respectively. Column four contains the X values and column five has the Y's. Unlike the prior example, note that X varies from row to row of the same subject. In this design, it is necessary to obtain a covariate for the between-subjects terms, (here A and S(A) and a different covariate for the within-subjects effects, (here B, AB, and SB(A)). Column six has been created by averaging the two X values for a given subject. Column six will then be used in exactly the same manner the X vector was used in the previous example, i.e., as the between subjects predictor variable. We shall label it here as SX to indicate it has just one value per subject.

Insert Table 6 about here

Column seven is created by subtracting column six from column four. It is labeled as the WX vector to indicate it is the regression vector to be used on within subjects effects. (The same results may be obtained by using column four as long as all models include column six prior to column four or other columns derived from column four. The use of column seven is desired only to improve the clarity of the example).

Vectors 8, 10, 12, 14, and 16 are created to reflect the orthogonal contrasts in the design. B is in column eight, the A effects are in columns 10 and 12, and the AB interaction in columns 14 and 16. Use of a unity vector and these vectors (plus subject vectors) in a multiple regression will yield the

Table 6

Linear Model Vectors of Winer's

[RS_3 in (A_3)] x B_2 w X Data

Columns

A	S	B	X	Y	SX	WX	ψ_B	$X\psi_B$	ψ_{A_1}	$SX\psi_{A_1}$	ψ_{A_2}	$SX\psi_{A_2}$	ψ_{AB_1}	$X\psi_{AB_1}$	ψ_{AB_2}	$X\psi_{AB_2}$
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	1	3.00	8.00	3.50	-.5	-1	-3.00	-2	-7.00	0	0	+2	6.00	0	0
1	1	2	4.00	14.0	3.50	+.5	+1	4.00	-2	-7.00	0	0	-2	-8.00	0	0
1	2	1	5.00	11.0	7.00	-2	-1	-5.00	-2	-14.0	0	0	+2	10.0	0	0
1	2	2	9.00	18.0	7.00	+2	+1	9.00	-2	-14.0	0	0	-2	-18.0	0	0
1	3	1	11.0	16.0	12.5	-1.5	-1	-11.0	-2	-25.0	0	0	+2	22.0	0	0
1	3	2	14.0	22.0	12.5	+1.5	+1	14.0	-2	-25.0	0	0	-2	-28.0	0	0
2	1	1	2.00	6.00	1.50	+.5	-1	-2.00	1	1.50	-1	-1.50	-1	-2.00	+1	2.00
2	1	2	1.00	8.00	1.50	-.5	+1	1.00	1	1.50	-1	-1.50	+1	1.00	-1	-1.00
2	2	1	8.00	12.0	8.50	-.5	-1	-8.00	1	8.50	-1	-8.50	-1	-8.00	+1	8.00
2	2	2	9.00	14.0	8.50	+.5	+1	9.00	1	8.50	-1	-8.50	+1	9.00	-1	-9.00
2	3	1	10.0	9.00	9.50	+.5	-1	-10.0	1	9.50	-1	-9.50	-1	-10.0	+1	10.0
2	3	2	9.00	10.0	9.50	-.5	+1	9.00	1	9.50	-1	-9.50	+1	9.00	-1	-9.00

Table 6 (cont.)

3	1	1	7.00	10.0	5.50	+1.5	-1	-7.00	1	5.50	+1	5.50	-1	-7.00	-1	-7.00
3	1	2	4.00	10.0	5.50	-1.5	+1	4.00	1	5.50	+1	5.50	+1	4.00	+1	4.00
3	2	1	8.00	14.0	9.00	-1	-1	-8.00	1	9.00	+1	9.00	-1	-8.00	-1	-8.00
3	2	2	10.0	18.0	9.00	+1	+1	10.0	1	9.00	+1	9.00	+1	10.0	+1	10.0
3	3	1	9.00	15.0	10.5	-1.5	-1	-9.00	1	10.5	+1	10.5	-1	-9.00	-1	-9.00
3	3	2	12.0	22.0	10.5	+1.5	+1	12.0	1	10.5	+1	10.5	+1	12.0	+1	12.0

AOV summary table reported as Table 10.6.7 (ii) in Winer (1971, p. 807).

A. Tests for Heterogeneous Slopes

Vectors 6 and 7 would be added as predictors to yield a COV summary table. However, prior to doing this, it is desirable to test for homogeneity of slopes. There are two types of regression slopes present in such a design, and they must be tested separately since they will have different error terms. Heterogeneity of between-subjects regressions (in the three A groups) is tested using the adjusted $MS_{S(A)}$. Heterogeneity of slopes on the within subjects factors is tested using adjusted $MS_{SB(A)}$. To provide a test of heterogeneous slopes on the A groups, multiply the SX vector by the A contrasts, yielding columns 11 and 13. To provide for a test of heterogeneous slopes on the within subjects factors, we multiply column seven by the B contrast and the AB contrast vectors. This yields columns 9, 15, and 17. Using a unity vector, and vectors 6 to 17 of Table 6 in one or more computer runs (depending on the programs available) in a process similar to that illustrated in the prior example yields the summary table of Table 7.

Insert Table 7 about here

The SS in the SX and WX rows are the sum of squares of regression associated with columns six and seven respectively. (These would not be included in some program outputs.) They clearly indicate the effectiveness of both covariants. The main interest in this model is whether we may assume homogeneity of slopes of the column seven vector over the several within subjects effects.

The between subjects homogeneity is tested by $MS_{Het\ A}/MS_{S(A)} = 2.528$. This value has a probability greater than .20, which leads to a retention of homogeneous slopes of the SX vector on the three A groups.

Table 7

Summary Table of the Completely Heterogeneous Slopes

Model Using Vectors 6 to 17 of Table 6

Source	DF	Sum of Squares	Mean Squares	F	P
BTN Subjects					
SX reg.	1	178.37			
A	2	54.259	27.129	4.926	>.10
Het A (cols. 11 & 13)	2	27.849	13.924	2.528	>.20
S(A)	3	16.522	5.5072		
Within Subjects					
WX reg.	1	62.745			
B	1	29.418	29.418	126.551	<.01
Het B (col. 9)	1	1.1179	1.1179	4.804	>.10
AB	2	1.9228	0.96142	4.136	>.10
Het AB (cols. 16 & 17)	2	1.8317	0.91584	3.940	>.10
Pooled 9, 16 & 17	3	2.9496	0.9832	4.230	>.10
SB(A)	2	0.46492	0.23246		
Total (ADJ)	17	374.50			

There are two different possible heterogeneous slopes tests for the within subject vector, WX. Using column nine we may test for heterogeneous slopes over the B main effect, while using columns 16 and 17 we may test for heterogeneous slopes over the AB interaction contrasts. Unless there is an apriori reason for believing that these would differ, these would be combined into a single test of heterogeneous slopes for the WX vector over all within subject effects. This is shown in Table 7 in the row with the pooled SS from vectors 9, 16, and 17. This test has a probability greater than .10 (as do the individual tests) so again the null hypothesis of homogeneity of slopes is retained. If the pooled test leads to the conclusion of heterogenity, it would be appropriate to test for homogeneity on B, and on AB separately.

B. Covariance Models

In the absence of heterogeneity, the covariance (homogeneous slopes) model is appropriate. This is given in Table 8. The table differs slightly from Winer's (1971, p. 807, iii) because Winer chose to use the regression coefficient for the WX vector also as the regression coefficient for the SX vector (they were quite close). The present solution uses the two vectors separately. Again the grand mean of the X's (7.5) is used to find the "adjusted means" as the predicted values of Y. In this case, the A main means, the B main means, and the cell means all will have changed somewhat due to the "adjustment" by covariance, since this included regression by both a between-subjects vector and a within-subjects vector.

Insert Table 8 about here

It is possible in some examples that only one of these two covariate vectors (column six and column seven) would have a significant regression, so you may wish to go to a simpler model with only one of the two. Since the

Table 8

Summary Table of the Covariance Model Using Vectors 6-8, 10, 12, & 14 of Table 6 and the Adjusted Means

Source	DF	Sum of Squares	Mean Squares	F	P
BTN Subjects					
SX reg.	1	178.37			
A	2	54.259	27.129	3.057	>.10
S (A)	5	44.370	8.8741		
Within Subjects					
WX reg.	1	62.745			
B	1	29.418	29.418	49.063	<.001
AB	2	2.3393	1.1696	1.9506	>.20
SB (A)	5	2.9980	0.59960		
Total (ADJ)	17	374.50			

Adjusted Means, $\bar{Y}'_{.jk}$

	B_1	B_2	$\bar{Y}'_{.j.}$
A_1	12.516	16.595	14.556
A_2	10.526	12.474	11.500
A_3	11.893	14.996	13.444
$\bar{Y}'_{..k}$	11.645	14.688	13.166

previous section illustrated the use of only the between subject vector, we shall illustrate an example where only the WX vector is retained, and the SX vector is dropped. Table 9 contains this summary table, and the adjusted means corresponding to it. The between subjects SS's are changed by dropping the SX vector, but we see that this has no effect on the within subjects SS's since the SX vector is orthogonal to all within subjects contrasts. For the same reason, the Btn-subjects SS's are the same as in the AOV table, and the main A means (a Btn-subjects effect) are the same as in the ANOVA analysis ignoring column seven. With the WX vector, as a lone covariate, the SS's within-subjects and the B main means are identical to what they were in the complete covariance analysis of Table 8. The cell means have been "adjusted" with respect to the WX covariate also. The general point is that in such an analysis of covariance you are always working with two orthogonal sets of effects. The between-subjects effects require one covariate, column six, and will be changed by it, but the between-subjects effects are orthogonal to the within-subjects effects, and the within-subjects covariate, column seven. Correspondingly, the within-subject effects are orthogonal to the between-subjects effects, and column six, the between-subjects covariate. Neither set is influenced by the decisions made on the other set.

This kind of covariance can be conceptualized as doing two different covariate analyses, one for the between-subjects effects, and one for the within-subjects effects. In each case, we should start by testing for heterogeneity of slopes, and proceed with the covariance only if the condition of homogeneous slopes is feasible. It is perfectly possible to have heterogeneous slopes on the between-subjects portion, and homogeneous slopes on the within-subjects effects; or vice-versa.

Insert Table 9 about here

VI. CONCLUSIONS AND SUMMARY

When working with organismic variables, one must face the fact that clear interpretative conclusions are not as easy to come by as when working with manipulative variables. The investigator must recognize that the organismic variables he has used are correlated with many other organismic variables, and it is impossible to be 100% confident that it is "rigidity" rather than some other variable correlated with rigidity that has produced the observed differences. The limitations of cross sectional studies can be viewed as the problem that the observed organismic variable, age, is confounded with other variables of educational differences, historical differences, environmental differences, etc., of cohorts. This same problem is present, in a different and lesser extent, when an organismic variable, X, is used in a longitudinal study.

Although covariance can be used as one method to try to eliminate some of the possible alternative interpretations, it is often used incautiously without testing even the basic assumption of homogeneous regression slopes. Least squares analyses are possible for either heterogeneous or homogeneous slopes models, with or without repeated measures. Since longitudinal studies require at least one repeated measure, such analyses have been illustrated in the present article. Unfortunately, even if all statistical assumptions have been met, it is still a matter of considerable controversy whether clear interpretations after a covariance analysis are justified if the $\bar{X}$'s are considerably different. Evans and Anastasio (1968), Ferguson (1966), and McNemar (1969) say yes, but Cronbach and Furby (1970), Campbell and Erlebacher (1970) and Games (1976) say no. The author thus recommends caution in such interpretations, and a willingness to look at the data from alternative interpretations.

Table 9

Analysis of Covariance with Only The WX Vector as a Covariate

Source	DF	Sum of Squares	Mean Squares	F	P
BTN Subjects					
A	2	100.00	50.000	1.695	>.20
S(A)	6	177.00	29.500		
Within Subjects					
WX reg.	1	62.745			
B	1	29.418	29.418	49.063	<.001
AB	2	2.3393	1.1696	1.951	>.20
SB (A)	5	2.9980	0.59960		
Total (ADJ)	17	374.50			

Adjusted Means, $\bar{Y}'_{.jk}$

	B_1	B_2	$\bar{Y}'_{.j.}$
	12.794	16.873	14.833
	8.859	10.808	9.833
	13.282	16.385	14.833
$\bar{Y}'_{..k}$	11.645	14.688	13.166

Most longitudinal studies are investigations, as distinguished from experiments (Games & Klare, p. 442). We must recognize we are mere observers in situations in which nature pulls a thousand strings. The strings we are watching may not be the crucial ones. Only by careful observation over many studies and many situations are the crucial strings likely to be identified, and only after they are identified are we likely to learn the crucial cues that indicate a desired response shortly will follow. The longitudinal investigator needs considerable patience, not only for the collection of his data, but for the processes to yield clear interpretations.

Univariate and Multivariate Analysis of Variance of Time-Structured Data

by

R. Darrell Bock
The University of Chicago

Table of Contents

UNIVARIATE AND MULTIVARIATE ANALYSIS OF VARIANCE OF TIME-STRUCTURED DATA

R. Darrell Bock
The University of Chicago

ABSTRACT

I. ASSUMPTIONS AND LIMITATIONS

II. TYPES OF TIME-STRUCTURED DATA

III. ANALYSIS OF CROSS-SECTIONAL DATA

IV. ANALYSIS OF LONGITUDINAL DATA

V. SUMMARY

Acknowledgement

Preparation of this paper was supported in part by NSF Grant BNS76-02849.

ABSTRACT

Detecting and describing change over time is a widely encountered methodological problem which, given certain design and measurement restrictions, can be effectively solved by analysis of variance and its multivariate extensions. Polynomial trend analysis is used to investigate the shape of the curve describing time-dependency of a population mean and differences in shape between populations, when data are cross sectional in nature. For longitudinal data, analysis of population time-point means or differences in trend between populations can be carried out by multivariate repeated measures analysis or, in favorable cases, by mixed-model univariate analysis of variance. Special data characteristics are discussed which lead to very powerful applications of the general models. Examples are presented and discussed in which the models are applied to physical growth data.

UNIVARIATE AND MULTIVARIATE ANALYSIS OF VARIANCE OF TIME-STRUCTURED DATA

Data are time structured when the observations can be identified with a number of pre-assigned points on the time continuum. In behavioral studies, we have the option of introducing time structure on more than one level: to study secular change in populations, we locate the observations in historical time (years, decades, centuries); to describe the growth and development of individual subjects within populations, we typically observe the subject at fixed intervals (days, months, years) on a time scale originating at his conception or birth; to characterize time-dependent response processes within subjects, we can record the subject's responses during intervals beginning at various elapsed times (seconds, minutes, hours) after the onset of an experimentally imposed condition or stimulus.

As a general technique for the statistical treatment of time-structured data, analysis of variance is virtually unique in its capacity to detect and summarize systematic time-dependent variation and covariation in observations from all of these levels, separately or jointly. In one unified analysis, it can encompass the possibly multiple outcomes of an experiment represented at several points in historical time, based on responses from subjects at different stages of development, and repeated within each subject on a number of occasions or trials. Given this structure, we could with the aid of analysis of variance and its multivariate extensions, extract the shape of the average response curve as a function of trial times, test for differences in shape due to the conditions imposed by the experimenter, to the stage of development of the subjects, or to secular change, and investigate all possible interactions of these factors.

Even in such complex applications the analysis of variance approach is surprisingly modest in its demands on mental effort to formulate the problem and on computing resources to perform the calculations. This conceptual and computational economy is not, however, purchased without a certain price. Because analysis of variance is part of linear least squares (Gauss-Markov) estimation, it requires simplifying limitations and assumptions for a straightforward and exact analysis. To benefit from the advantages, we have to be cognizant of the limitations. We must not attempt to extend this general purpose technique into realms where more specialized methods (often involving non-linear estimation) are required. Obviously, we must begin our investigation with the plan of the data analysis and its restrictions clearly in view. Some of the points to be considered at the planning stage are discussed in the next section.

I. ASSUMPTIONS AND LIMITATIONS

For time-structured data to be accessible to analysis of variance, not only must each observation carry a time identification, but the method of measuring response, the sampling plan, the arrangement of the time points, the assumed form of time-dependency, and the nature of the error distribution must be restricted in certain ways.

A. Interval Measurement

If the object of the analysis is to describe systematic trend over time, it is mandatory that the response variables be measured on a scale whose units are commensurate throughout the relevant range of variation--i.e., the measurement must be on a so-called "interval" scale with units everywhere of constant size in some well-defined sense. Otherwise, the shape of the trend line is arbitrary, straight lines cannot be distinguished from curves, and parallel lines at different elevations may appear non-parallel. Even when the measures are known to be equal on one scale (e.g., physical units such as gram, centimeter, seconds), it may be desirable to transform them to equality on another scale in order to see clearly the essential form of the trend line. For example, variables measuring exponential growth may be better analyzed in log units rather than the original units.

The problem of commensurate units can be especially severe when the measures arise from behavioral responses. Bock and Jones (1968, Chapter 1) discuss the issue of defining measurement scales for behavioral data that have some of the properties of, for example, the c.g.s. system in physical measurement. They point out that physical units are intrinsically defined not by the method by which the measurements are taken, but by their role in the mathematical models that connect one observable phenomenon with another. Thus, the units of length take on meaning when the formula for computing area can be used at every point on the scale of measurement, and that area can be used to translate pressure into force in terms of units of mass, and so on. Specifically, it is the invariance of a great variety of mathematical models with respect to location on the measurement scale that gives meaning, utility and generality to systems such as c.g.s.

With admittedly weak theoretical underpinning, we can perhaps accept as having units any behavioral measurement scale that has a validated linear relationship with another variable of interest. By this criterion we might be willing to accept the Binet I.Q. scale as interval measurement on the grounds that in the interval from five to fifteen years it exhibits a linear relationship with many other physiological and psychological indices of maturation. To the extent that Binet I.Q. differences translate proportionately into increases in these measures, the assumption that the I.Q. scale has well defined units is not entirely gratuitous.

But stronger definitions of behavioral scales are possible. Bock and

Jones (1968) consider Thurstone's psychological scaling to be interval measurement defined by related response models. They present a number of models for judgment and choice that are connected by a common scale and, using units on this scale, perform analyses of variance of factorial models for sensation and preference. In much the same sense, modern psychometric methods of latent trait measurement produce scales with commensurate units by defining a model relating differences on the scale to item-response probabilities (Bock, 1972; Lord, 1974; Rasch, 1960; Samejima, 1969). These scales appear also to yield linear relationships with other variables (Andersen, 1976; Bock, 1976; Bock and Thrash, 1976). Because much of research on human behavior at the individual level depends on objective test instruments, it is of considerable interest that latent trait theory can open this domain to statistical methods, such as analysis of variance, which assume interval measurement.

B. Group Comparisons

If a straightforward application of analysis of variance is desired, only estimation or comparison of group means should be considered. The questions the investigator can ask of the data are limited to those concerning the shape of the curve of population means as a function of time, or differences between the means of two or more populations as a function of time. Whether or not these are interesting questions depends critically upon the practical meaningfulness of group averages. It has long been recognized that such averages are not completely informative about individual development. A well-known example is the unsatisfactory characterization of the adolescent growth spurt in plots of mean stature versus age. The spurt is apparent in such data, but few if any subjects follow the mean curve in their own growth. The group mean curve tends to show a more gentle spurt because of the averaging of individual growth spurts occurring at different times. But even this generalization is not entirely true because there are some subjects, especially among the boys, who show a more gentle and protracted slope than is seen in the average data.

The positive thing that can be said of average data is that, if there is some weak but consistent trend in the population, the average growth curve may detect it whereas the examination of individual growth curves separately may not. An example of this phenomenon presented in Bock (1976) suggests that, without resort to advanced treatment of individual growth curves including non-linear models and Bayes estimation (see Bock, 1976), the investigator will be limited to characterization and comparison trends of group averages.

C. Fixed Time Points

Analysis of variance can be applied conveniently to time-dependent data only when the time points are fixed in advance and are moderate in number. In studies where chronological age is the time variable, this requirement can be met by measuring a subject at pre-selected ages. In growth studies such as the Fels or Berkeley studies, for example, the children were measured on or near their birth dates or, at younger ages, at their year and half-year anniversaries. If this degree of pre-planning is not possible, the subjects will have a more or less random distribution of age at the time of measurement. For purposes of the analysis of variance, the data must then be grouped into age ranges and the mean or median age used to represent the group. In educational work, similar use of grade-in-school as a time point may be defended on grounds that the relevant dimension for growth of achievement is years of schooling rather than chronological age. Although not a mandatory requirement, it is also convenient for the time points to be evenly spaced. As we shall see, the analysis of variance of trend is then more easily carried out.

D. Short-term Moderate Change

For a number of reasons, analysis of variance techniques may become difficult to apply when there are many time points encompassing substantial change in the variables of interest. This is not simply a matter of the computational labor in analyzing designs with many time points, but one of increasing difficulty in justifying the assumptions of conventional trend analysis. As discussed in the next section, we typically use a low-degree polynomial to represent group mean curves and differences in mean curves. This is quite satisfactory for short-run change, but may not be suitable over a wider range. The curve for average growth in stature, for example, has no very satisfactory polynomial representation over the entire growth cycle. Only non-linear models such as the two-component logistic model of Bock, et al. (1973), or the three-component model of Bock and Thissen (1976), seem capable of describing growth in stature from near birth to maturity. Fitting and testing of these models requires non-linear estimation and cannot be approached by the elementary methods of univariate or multivariate analysis of variance discussed here. However, growth over a more limited range can be so described, as will be apparent in the examples in Sections III and IV.

E. Freedom From Outliers

Like all least-squares techniques, analysis of variance is adversely affected by a few aberrant observations far removed from the main body of the data. In behavioral and biological measurement, such aberrant values are almost always the result of clerical errors, or of subjects in the sample who do not actually belong to the populations sampled. Fortunately, outliers are easy to detect when pre-screening the data and can be removed from the sample before the analysis begins.

To justify linear least-square estimation in terms of unbiasedness and minimum variance, it is only necessary to assume that the error distribution has finite mean and finite constant variance throughout the range of measurement. (The presence of outliers indicates that the assumption of homogeneous variance has been violated.) To justify the nominal error rates of the significance tests associated with analysis of variance, it is necessary to add the assumption of normally distributed error or large-sample assumptions. Because many of the measures used in behavioral studies are essentially additive combinations of many more or less independent sources of environmental, biological and physiological variation, the assumption that a normal error distribution, after systematic effects have been absorbed into the model, is broadly justified. The only notable exceptions are response-time measures, which tend to have a log-normal rather than a normal distribution. In many applications, the logarithms of response times are satisfactory quantities for analysis of variance (see Thissen, 1976).

II. TYPES OF TIME-STRUCTURED DATA

Insofar as it affects the method of analysis, the main distinction to be made is between cross-sectional and longitudinal data.

A. Cross-sectional Data

In cross-sectional data, different subjects are sampled at each time point and all measures in the sample are assumed to be statistically independent. As a result, in the crossed design of (experimental or sampling) groups x time points, the observations are independent both within and between cells. They may therefore be analyzed in a conventional two-way or multi-way analysis of variance with but one feature particularly related to time dependence--that in the partition of the sum of squares for the time way-of-classification and its interactions, single-degree-of-freedom terms are isolated for each component of trend.

If proportionate numbers of subjects in each group appear in each time-point class, a straightforward orthogonal analysis of variance applies. If the numbers are disproportionate, a non-orthogonal analysis will be necessary and will require the investigator to fix the order of the partition of sum of squares by choosing an order of priority among hypotheses about various effects in the model. This issue will be clarified in the discussion of analysis of variance of cross-sectional data illustrated by an example from anthropology, presented in Section III.

B. Longitudinal Data

Time-structured data are longitudinal when each subject is measured on a scale commensurate at each time point. Note that, while a longitudinal study resulting in this type of data is prospective, not all prospective studies are longitudinal or even time-structured. A study that obtains one set of measures at an earlier time, and a second qualitatively distinct set at some later time, is prospective and may enable prediction of later characteristics from earlier. But it is not longitudinal or time-structured, does not describe change or growth over some period of time, and cannot be subjected to analysis of variance. Longitudinal studies are both prospective and time-structured, enable both prediction and description of growth and change, and are amenable to analysis of variance.

The analysis of variance of longitudinal data is more complex and interesting than that of cross-sectional data. In the psychological and behavioral literature, the statistical treatment of longitudinal data is often called "repeated measures analysis" (Bock, 1975, Chapter 7; Winer, 1971). In the biometric and statistical literature, this topic is usually referred to as "analysis of growth" or of "growth curves" (Khatri, 1966; Lee, 1974; Pottoff & Roy, 1964). Basically, three forms of repeated measures analysis have been proposed:

1) Mixed-model univariate analysis of variance (Lindquist, 1953; Winer, 1971).
2) Unweighted (exact) multivariate analysis of variance (Bock, 1963).
3) Weighted (large-sample) multivariate analysis of variance (Khatri, 1966; Pottoff & Roy, 1964).

The choice among these methods depends upon the nature of the time-dependency and on the structure of the variance-covariance matrix of the residuals from the fitted trend line. In most cases, this choice can be made only after some preliminary inspection of the data. How this inspection is carried out and the subsequent analysis performed is discussed and illustrated in Section IV.

C. Other Characteristics of Longitudinal Data

A sometimes troublesome limitation on the multivariate analysis of longitudinal data is that the data for each subject must be complete. Although recent work on the problem of multivariate-wise incomplete data appears promising (Kleinbaum, 1973; Rubin, 1974; Trawinski & Bargmann, 1964), the practical implementation of these developments is still several years off.

In the meantime, the investigator faced with missing data has the option of (1) omitting subjects with incomplete data records (if this makes the experimental or sampling design unbalanced, a non-orthogonal multivariate analysis of variance will be required), (2) proceeding under mixed-model assumptions with a non-orthogonal univariate analysis of variance (because subjects must be included as a way-of-classification in this analysis, the computations will be extremely heavy if the number of subjects is large), (3) using some method of interpolating data points if relatively few records are incomplete. Considering the problems that may attend any of these options, the investigator is better advised to expend his energy on collecting complete data initially than to attempt a patch-up later.

As a preface to the discussion of statistical methods, it should perhaps be mentioned that, except in the simplest cases, the computations will require the use of a fairly large-scale computer program. Of the several multivariate analysis of variance programs available, the MULTIVARIANCE program of Jeremy Finn (1974) is the most convenient for repeated measures analysis, and Version VI, which is in progress, greatly extends this facility (Finn, 1976). (The examples in the present paper were prepared with Version V.) A discussion of the use of the MULTIVARIANCE program in the analysis of time-structured data appears in Finn and Mattsson (1977).

III. ANALYSIS OF CROSS-SECTIONAL DATA[1]

In studies of human growth and development, limited time and resources may leave the investigator no option except to collect cross-sectional data. A good example are the measures of height and weight, collected by Haller, et al. (1967) and more recently by Jamison (1977), among children indigenous to the Alaskan North Slope. Although purely cross-sectional, these data are quite adequate to check on anecdotal reports that these children grow more slowly than children living in the south 48 states. An analysis of some of Jamison's results in comparison with a control group of children the same age is presented in the computing example at the end of this section.

A. Form of Cross-sectional Data

Cross-sectional data may be represented in the form of a so-called "crossed" analysis of variance design (groups x occasions) shown in Table 1. Note that the necessary limitation to time points identical for all groups, is represented in Table 1 by the recurring value of x_k. Note also that the numbers N_{jk} of independently sampled subjects in the group x occasion subclasses are not necessarily assumed equal--indeed, a non-orthogonal analysis of these data is still possible even when some of the N_{jk} are zero.

Insert Table 1 about here

The sample statistics required for least-squares analysis of this form of data are the subclass means $y_{\cdot jk} = \sum_{i=1}^{N_{jk}} y_{ijk}/N_{jk}$,

the subclass numbers N_{jk} and the pooled within-subclass variance estimate,

$$\hat{\sigma}_N^2 = [\sum_{j=1}^{n} \sum_{k=1}^{m} \sum_{i=1}^{N_{jk}} (y_{ijk}^2 - N_{jk} y_{\cdot jk}^2)]/(N-n) \quad ,$$

TABLE 1

Form of Time-Structured Cross-sectional Data

Groups j	Time-points x_k	Observations y_{ijk} $(i=1,2,\ldots,N_{jk})$
1	x_1	y_{i11}
	x_2	y_{i12}
	$\vdots$	$\vdots$
	x_m	y_{i1m}
2	x_1	y_{i21}
	x_2	y_{i22}
	$\vdots$	$\vdots$
	x_m	y_{i2m}
$\vdots$	$\vdots$	$\vdots$
n	x_1	y_{in1}
	x_2	y_{in2}
	$\vdots$	$\vdots$
	x_m	y_{inm}

where $N = \sum_{j=1}^{n} \sum_{k=1}^{m} N_{jk}$.

The first objective of the data analysis is to choose the linear model that describes systematic effects in the observations. The criterion for this choice is one of plausibility and parsimony--namely, that the model should comprise the least number of effects that accord with theory concerning the phenomenon in question and with acceptable fit of the model as judged by a formal goodness-of-fit test.

The second objective is, given the data, to estimate the effects and their standard errors, to compute from the estimated effects the expected trend lines for the groups, and to show the expected dispersion of observations about the trend line, possibly in the form of a tolerance interval for a new observation at given time points. The first step in reaching either of these objectives is to formulate a class of models for time trend.

B. The Polynomial Model For Time Trend

If some interval-measured time-dependent variable y has been observed on N_k subjects at successive distinct fixed time points x_k, k=1, 2,...,m, and m is not too large nor the change in y discontinuous in this interval, a suitable statistical model for time trend is the $q<m$ degree polynomial with additive error,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_q x^q + \varepsilon \ . \qquad \text{(III. 1)}$$

The random error, ε, is assumed to be independently distributed with mean 0 and unknown variance, σ^2. The β's are in general unknown, but may be estimated from the means of the observations at each time point, $y_{.k} = \frac{1}{N_j} \sum_{i=1}^{N_k} y_{ik}$, by the so-called Gauss-Markov (least-squares) estimator,

$$\underline{\hat{\beta}} = (X'DX)^{-1} X'D\underline{y}_{.} \ , \quad \text{for} \quad |X'DX| \neq 0 \ , \qquad \text{(III. 2)}$$

where $\underline{\hat{\beta}}' = [\beta_0, \beta_1, \beta_2, \ldots, \beta_q]$,

$\underline{y}.' = [y._1, y._2, \ldots, y._m]$,

$D = \mathrm{diag}\,[N_1, N_2, \ldots, N_m]$,

and

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^q \\ 1 & x_2 & x_2^2 & \cdots & x_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^q \end{bmatrix} . \qquad \text{(III.3)}$$

The m×(q+1) matrix X contains the leading q+1 rows of the Vandermonde matrix of order m. Since X is of rank q+1 when all x are distinct (Browne, 1958), $|X'DX| \neq 0$ in general. The expected value of this (unbiased, minimum-variance linear) estimator is $\underline{\beta}$ and its sampling variance-covariance matrix is $\sigma^2(X'DX)^{-1}$ (see MSMBR, Sec. 4.1).

Although satisfactory in most other respects, this method of fitting the polynomial model by estimating the β's has the disadvantage of requiring the degree q to be specified in advance. In general, all elements of $(X'DX)^{-1}$ and of $\underline{\hat{\beta}}$ change when columns are added to or deleted from X. Yet in many cases the investigator is uncertain about the least-degree that will give a good account of the data and will wish to inspect the goodness-of-fit of several successively lower-degree models before deciding.

To facilitate this form of stepwise testing of polynomial models and to make the calculations easier, Fisher (1921) introduced a method of reparameterizing (III.1) as a so-called orthogonal polynomial model. The reparameterization is equivalent to applying to X the Gram-Schmidt orthogonalization with respect to the metric matrix D, beginning in the left-most column of X and working to the right (see MSMBR, Sec. 2.2.4). The result is to decompose X into an m×(g+1) orthogonal matrix P and an upper triangular matrix S'. This is,

$$X = PS' , \qquad (III.4)$$

where $P'DP = I$ and $X'DX = SS'$.

Then (III.1) may be expressed in terms of the orthogonal coefficients $\underline{\gamma} = S'\underline{\beta}$, which are estimated from the group means by

$$\hat{\underline{\gamma}} = P'D\underline{y}_{\cdot} = \underline{u} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_q{}^1 \end{bmatrix} \qquad (III.5)$$

The elements of the (g+1)x 1 vector $\underline{u}$ have expected value $S'\underline{\beta}$; their variance-covariance matrix is the (g+1) x (g+1) identity matrix--that is, they have unit variance and are uncorrelated. The convenience of this parameterization is due to the implied statistical independence of the orthogonal estimates when the observations are normally distributed and to the fact that S' is upper triangular with strictly positive diagonal elements. The former property implies that, on the hypothesis that the corresponding orthogonal coefficient is null, the square of each element in $\underline{u}$ is distributed independently as a central chi-square on one degree of freedom. The latter implies that accepting the null hypothesis for the last q_2 orthogonal coefficients is equivalent to accepting $q_1 = q-q_2$ for the degree of the polynomial model for trend. Together, they justify the averaging of squares of the last q_2 elements of $\underline{u}$ for use as the numerator mean-square of an F statistic testing goodness-of-fit of the degree q_1 polynomial vs. the degree q polynomial. The denominator is either the within-group mean-square or the residual mean-square for the degree q model obtained by pooling the residual between-group sum of squares and the within-group sum of squares. The calculations involved in this test are summarized in Table 2. If on the basis of this test, the degree q_1 model

---

Insert Table 2 about here

---

is adopted, the q_1 +1 estimated coefficients of the polynomial are given by

$$\hat{\underline{\beta}}_1 = (S'_{11})^{-1}\underline{u}_1 \quad , \qquad \text{(III.6)}$$

where S_{11} is the leading $(q_1+1)x(q_1+1)$ submatrix of S, and $\underline{u}_1$ contains the leading q+1 elements of $\underline{u}$. The variance-covariance matrix of this estimator is $\sigma^2 (S'_{11})^{-1} S_{11}^{-1}$. The error variance, σ^2, is estimated by the denominator mean square in the goodness-of-fit statistic.

But it is not always necessary to compute the estimated β- coefficients. In most cases the trend line can be plotted from the fitted values at the assigned time points, and these may be computed from the orthogonal estimates by

$$\hat{\underline{y}} = P_1\underline{u}_1 , \qquad \text{(III.7)}$$

where P_1 contains the leading q_1+1 columns of P. Similarly, the 2σ tolerance interval for a new observation at the point x_j is

$$\hat{y}_{.j} \pm 2\sigma\sqrt{1 + [P_1]'_j \; [P_1]_j} \quad , \qquad \text{(III.8)}$$

where $[P_1]_j$ is the j-th row of P_1 written as a column. (See MSMBR, Sec. 4.1.13.)

B. The Fisher-Tchebycheff Orthogonal Polynomials

The true merit of Fisher's method of fitting the polynomial model for trend is most evident when the time points are spaced equally (i.e., when

TABLE 2

Cross-sectional Data: Analysis of Variance for Testing the Fit of a q_1 Degree Polynomial, Given that the q Degree Model has Been Found to Fit. The Regression Sums of Squares are Computed From the Orthogonal Estimates $\underline{u} = P' D\underline{y}$.

Source of Variation	Degrees of Freedom	Sums of Squares	F-statistic*
Constant	1	$ssm = u_o^2$	
Linear	1	u_1^2	
Quadratic	1		
⋮	⋮	⋮	
q_1-ic	1	$u_{q_1}^2$	
q_1 degree model	q_1	$ssr_1 = \sum_{\ell=1}^{q_1} u_\ell^2$	
q degree model, given q_1 degree	$q_2 = q-q_1$	$ssr_2 = \sum_{\ell=q_1+1}^{q} u_\ell^2$	$\frac{ssr_2/q_2}{(sse+ssw)/(N-q-1)}$
Between-groups residual	m-q	$sse = ssg-ssr_1-ssr_2-ssm$	
Group means	m	$ssg = \sum_{k=1}^{m} N_k y_{.k}^2$	
Within groups	N-m	ssw = sst-ssg	
Total	$N = \sum_{k=1}^{m} N_k$	$sst = \sum_{k=1}^{m} \sum_{i=1}^{N_k} y_{ik}^2$	

*or $(ssr_2/q_2)/[ssw/(N-m)]$

$x_{j+1} - x_j$ is constant for j=1,2,...,m-1). For then, P is invariant with respect to the origin and units of the time measure and is, in fact, precisely the tabled Fisher-Tchebycheff orthogonal polynomial. Provided the values of x belong to the rational numbers, the elements of P are rational and may be given in integer form with respect to a largest common denominator, as may the square of the normalizing constants for the columns of P. As a result, the orthogonal estimates can, for given data, be computed up to the limit of the tables without any rounding error whatsoever. Similarly, the matrix S may be tabled in integers and inverted in integer operations to obtain estimates of $\underline{\beta}$ without error. Thus, the problem of round-off error, which plagues least-squares fitting of high-order polynomials (Wampler, 1970) is completely solved if the orthogonal polynomials are used. The tables given in MSMBR, Appendix B, for polynomials up to q=9 include the integer forms of the P matrix, the normalizing constants, and the S matrix. They are convenient for othogonal polynomial trend analysis when the number of time points does not exceed 10. For greater numbers of points, the DeLury (1950) and Fisher-Yates (1963) tables are available.

C. Orthogonal Polynomial Trend Analysis in Cross-sectional Data

For cross-sectional designs in which, say, m time points are equally spaced, the orthogonal polynomial matrix of order m may be incorporated into the analysis of variance to provide up to m-1 single degree-of-freedom components of the sum of squares for the time dimension and for its interaction with the other ways of classification in the design. Except for the additional ways of classification of the groups, this analysis is similar to that shown in Table 1 for the time classification only. In the numerical illustration to follow, the subjects are classified by the location of the communities from which the subjects were sampled (which in this case reflect differences in the ethnic origin of the samples) and further classified by sex, and by years of age. Thus, the design may be described as a Location x Sex x Age cross-classification.

Using the conventional model for crossed designs (see MSMBR, Sec. 5.3), the parameter space of the model, and corresponding sums of squares in the analysis of variance is partitioned into the following subspaces: General mean, Location, Sex, Age, Location x Sex, Location x Age, Sex x Age, and Location x Sex x Age.

The purpose of the analysis of variance for this type of design is to aid in the choice of the least complex model for effects of the sample classes (in this case the Location and Sex groups) and the lowest degree model for polynomial trend in any of these effects. If there is a significant way of classification in the analysis, the corresponding parameter space is retained in the model. If there is a significant two-factor interaction involving a given way of classification, then the two-factor space and the main class space of that way of classification is retained. Similarly, if there is a significant three-factor interaction involving a given way of classification, the spaces corresponding to that interaction, to the two-factor interactions involving that classification and the main class space are retained. And so on, to the highest order of interaction.

In any of these spaces if a q_1 degree polynomial is required for a given way of classification in any main class or interaction space, then the q_1 degree polynomial is used in all spaces involving that way of classification when fitting the final model. Trend in the data is depicted by computing, from the fitted degree q_1 polynomial, the marginal or group means required for plotting interaction or main-class effects as appropriate. A plot of trend lines is included in the computing example.

If the design is non-orthogonal (i.e., the subclass numbers are disproportionate), the order in which the subspaces enter the model must be specified by the investigator. In general, the strategy is to enter effects with more prior certainty of existence before those with less prior certainty. This provides a critical test of the more dubious effects, unconfounded by effects that are presumed to exist and are necessary in the model.

In the case of the Location x Sex x Age design, age effects are a foregone conclusion, sex effects are always possible, but there is little prior knowledge about location effects. If interactions are considered less certain than main effects, a reasonable ordering of spaces for the analysis of variance might be:

Mean, Age, Sex, Location, Age x Sex, Age x Location,
Location x Sex, Age x Location x Sex.

When there is ambiguity as to the ordering, the analysis may be carried in more than one order. But such analyses are in general not independent and should be held to a minimum to avoid incurring Type I errors considerably more frequently than their nominal rates.

D. EXAMPLE 1: Cross-sectional comparison of growth in stature of children aged 6 to 14 years from two populations

To illustrate the analysis of variance of trend in cross-sectional data, we compare some data reported by Jamison (1977) giving the stature of boys and girls age 6 to 14 years from the villages of Barrow and Wainwright, Alaska, with data of Tuddenham and Snyder (1954, p. 199) for boys and girls in the Berkeley Guidance Study. Strictly speaking, this is not a rigorous analysis because the Berkeley data are actually longitudinal (and will be analyzed longitudinally in Example 2). But it clarifies the calculations and gives some indication of the results that might be expected from actual cross-sectional data.

Sample statistics required in the calculations are shown in Table 3. The within-age-group variance estimate is reconstructed from the age-group

----

Insert Table 3 about here

----

sample standard deviations reported in the original sources.

The calculations outlined in Section III were carried out on these data by means of the MULTIVARIANCE program (Finn, 1974). This program provides both the non-orthogonal analysis of variance and the orthogonal polynomial trend analysis required in this problem. After the terms to be retained in the model are chosen, the program computes the estimated orthogonal polynomial coefficients and the predicted values for the mean-trend lines for the groups.

In this application, the groups are cross-classified by location (North Slope, Berkeley) and by sex (male, female), and the between-group and group x occasion interaction degrees of freedom are partitioned accordingly in the analysis of variance shown in Table 4. Note also that degrees of freedom for polynomial trend of higher degree than quartic are pooled in this table.

Insert Table 4 about here

The results of the analysis of variance in Table 4 are clear enough, with one minor exception. There is some evidence of Location x Occasion trend effects of degree 5 through 8 (p=.046). Significant high-degree orthogonal polynomial components almost always indicate the presence of one or two irregular points in the data, often because of procedural or clerical errors. From the plot of the group means in Figure 1, it appears that age groups 10 and 13 are out of line both for boys and girls in Jamison's data. This is undoubtedly due to non-random sampling aggravated by the pooling of samples from several sites and by elevated coefficients of consanguinity in the North Slope villages.

TABLE 3

Mean Stature (cm.) of Boys and Girls Aged 6 Through 14: North Slope and Berkeley Samples[a]

	North Slope[b]				Berkeley[c]			
Age (Years)	Boys Mean	N	Girls Mean	N	Boys Mean	N	Girls Mean	N
6	113.0	22	113.4	11	117.5	66	117.2	70
7	117.8	15	118.2	22	124.0	66	123.4	70
8	122.4	15	122.5	19	130.1	66	129.2	70
9	129.9	20	129.2	22	135.9	66	135.2	70
10	132.2	23	130.1	16	141.3	66	141.0	70
11	137.9	14	137.6	18	146.5	66	147.6	70
12	144.1	26	147.2	19	152.2	66	154.5	70
13	145.1	14	148.1	16	158.8	66	159.8	70
14	154.0	13	155.1	12	165.8	66	163.1	70

[a] Common within-group standard deviation = 6.23

[b] Jamison (1977)

[c] Tuddenham and Snyder (1954, p. 199)

TABLE 4

Cross-sectional Trend Analysis of Average Stature of Children Aged 6 Through 14 from Berkeley, California, and the Alaskan North Slope (Age x Sex x Location)

Source of Variation	d.f.	Sum of Squares	F	p
General Mean	1	---	---	---
Linear Age	1	336,242	8666	<.0001
Quadratic Age	1	11.96	.30	.58
Cubic Age	1	4.48	.11	.74
Higher Age	5	98.91	.51	.77
Sex	1	6.11	.16	.69
Location	1	15,766	406	<.0001
Lin. Age x Sex	1	36.75	.95	.33
Quad. Age x Sex	1	73.55	1.90	.17
Cubic Age x Sex	1	319.40	8.23	.004
Higher Age x Sex	5	190.36	.98	.43
Lin. Age x Loc.	1	1,004.6	25.89	<.0001
Quad. Age x Loc.	1	18.75	.48	.49
Cubic Age x Loc.	1	.002	.000	.99
Higher Age x Loc.	5	437.35	2.25	.046
Loc. x Sex	1	21.25	.55	.46
Age x Loc. x Sex	5	140.97	.45	.89
Within Groups	1505	58,394		

Insert Figure 1 about here

If the significant higher-degree Age x Location interaction is discounted, the only significant greater-than-linear effect is the cubic Age x Sex interaction ($p=.004$). Because girls reach mature stature before boys, this type of interaction is to be expected as the children enter adolescence. It is clearly seen in the group means in Table 1 as an inversion in the order of the means for the two sexes in the Berkeley data. That there is no similar inversion in the North Slope data might suggest an Age x Sex x Location interaction, but the analysis does not confirm its presence.

If the cubic Age x Sex term is included in the model, the Sex main effect and Age main effects up to degree 3 must also be retained along with the highly significant Location effect and the linear Age x Location interaction. The latter confirms the reality of a difference in growth rate between the Berkeley and North Slope populations during the long period of essentially linear-in-age preadolescent growth in stature. From the fact that the Location contrast is North Slope minus Berkeley (A-B), and the linear Age x Location contrast is negative, we deduce that the Berkeley population is growing faster. This is confirmed by the plot in Figure 1, of the fitted group means calculated from the orthogonal estimates in Table 5. Between 8 and 14 years of age, the rate of growth of the Berkeley children is about .8 cm/year greater

----

Insert Table 5 about here

----

than that of the North Slope children. This figure is obtained by dividing the orthogonal estimate by the normalizing coefficient for the linear orthogonal polynomial of order 9:

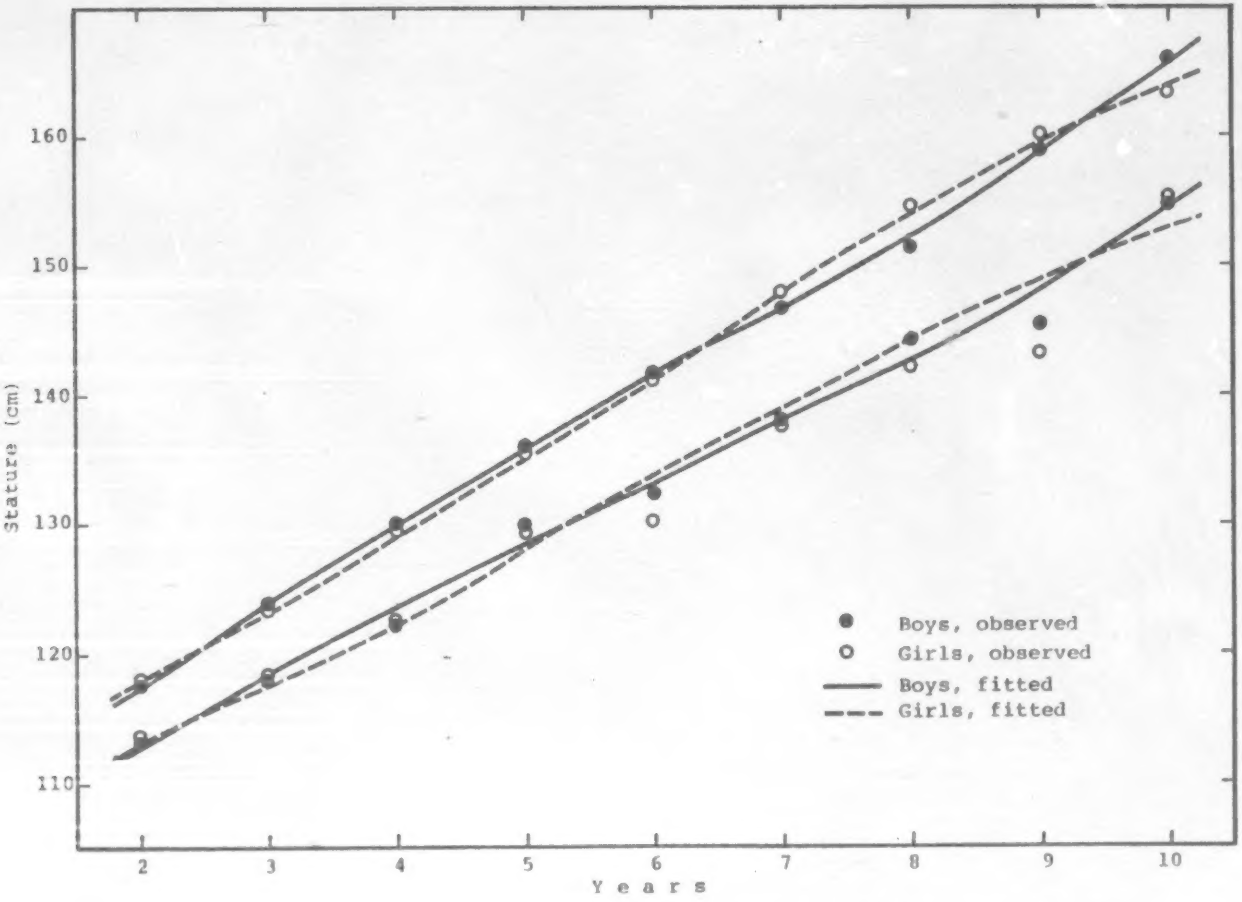


Fig. 1. Fitted growth curves for the Berkeley and Barrow samples

TABLE 5

Orthogonal Estimates: Rank 10 Polynomial Model

Effect	Estimate	S.E.
Constant	137.3027	.1968
Linear Age	42.5591	.6143
Quadratic Age	-0.2143	.4805
Cubic Age	-0.0620	.4782
Sex (Boys-Girls)	0.0096	.3176
Location (NS-B)	-7.9580	.3937
Lin. Age x Sex	-0.6483	.9625
Quad. Age x Sex	1.1824	.9603
Cubic Age x Sex	2.9167	.9567
Lin. Age x Loc.	-6.3038	1.2294

$$-6.3038/\sqrt{60} = -0.8138$$

(See MSMBR, Appendix B.)

IV. ANALYSIS OF LONGITUDINAL DATA

In a fully longitudinal study, each and every subject is measured at the same or equivalent, pre-assigned time points.[2] The formal layout of data from

----

Insert Table 6 about here

----

such a study is shown in Table 6. Note that the measure repeated on each subject is indexed by the superscript k=1,2,...,p (in parentheses to distinguish it from an exponent). The subscript is reserved for the identification of experimental or sampling group j and for subject i within group j. The (arbitrary) number of subjects in each group is N_j. If the groups are further classified according to experimental factors and/or sampling attributes, j may be replaced by a multiple subscript indicating the treatment or attribute combination.

A. Sample Statistics

All computations of a linear least-squares analysis of longitudinal data may be performed starting from the following summary information:

1) The group vector means

$$\underline{y}_{.j}' = [y_{.j}^{(1)} \; y_{.j}^{(2)} \ldots y_{.j}^{(p)}], \qquad \text{(IV.1)}$$

where

$$y_{.j}^{(k)} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij}^{(k)} \;.$$

TABLE 6

Form of Longitudinal Time-Structured Data

	j	x_1	x_2	...	x_p	
	1	$y_{i1}^{(1)}$	$y_{i1}^{(2)}$	...	$y_{i1}^{(p)}$;	$i = 1,2,\ldots,N_1$
Groups	2	$y_{i2}^{(1)}$	$y_{i2}^{(2)}$	...	$y_{i2}^{(p)}$;	$i = 1,2,\ldots,N_2$
	$\vdots$					
	n	$y_{in}^{(1)}$	$y_{in}^{(2)}$	...	$y_{in}^{(p)}$;	$i = 1,2,\ldots,N_n$

2) The subclass numbers N_j, $j=1,2,\ldots,n$.

3) The pooled within-group variance-covariance matrix $\hat{\Sigma}$ in which the diagonal elements are the unbiased variance estimates,

$$\hat{\sigma}_k^2 = \frac{1}{N-n} \{ \sum_{j=1}^{n} \sum_{i=1}^{N_j} [y_{ij}^{(k)}]^2 - \sum_{j=1}^{n} N_j [y_{.j}^{(k)}]^2 \} , \qquad \text{(IV.2)}$$

and the off-diagonal elements are the unbiased covariance estimates

$$\hat{\sigma}_{k\ell} = \frac{1}{N-n} \{ \sum_{j=1}^{n} \sum_{i=1}^{N_j} y_{ij}^{(k)} y_{ij}^{(\ell)} - \sum_{j=1}^{n} N_j y_{.j}^{(k)} y_{.j}^{(\ell)} \} , \quad \ell \neq k \qquad \text{(IV.3)}$$

where $N = \sum_{j=1}^{n} N_j$.

A typical example of longitudinal data summarized in this form is shown in Table 8. The upper section of Table 8 contains the mean yearly measures of stature (cm.) for boys and girls ages 2 through 8 in the Berkeley study as published by Tuddenham and Snyder (1954). The lower section contains the unbiased estimate of the common within-sex-group variation and covariation. Note that to facilitate visual inspection, the latter is shown in the form of standard deviations and correlations rather than variances and covariances.

B. The Polynomial Model for Longitudinal Data

A general linear model suitable for repeated measures data was first given by Roy (1957, p 83) and applied to the analysis of change in Bock (1963) and to the analysis of growth curves by Khatri (1966), Pottoff and Roy (1964), and numerous subsequent workers (Kleinbaum, 1973; Lee, 1974; MSMBR, Sec. 7.2; Timm, 1975, Sec. 5.16; Tubbs, Lewis and Duran, 1975). For present purposes, it is most convenient to express this model in terms of the nxp matrix Y. of vector means for the experimental or sampling groups:

$$Y. = A\beta X' + \xi \qquad \text{(IV. 4)}$$

In (12), the n x m design matrix A contains 1 and 0 elements specifying how the trend effects enter additively into the expected values of the group means.

The columns of the m x s matrix β pertain to the unknown coefficients of the degree s-1 polynomial model and the rows pertain to main class and possible interactive effects of the experimental factors or sampling attributes.

The p x s matrix X contains the leading $s \leq p$ columns of the order-p Vandermonde matrix shown in (III.3).

Row j of the n x p matrix $\xi.$ contains the means of errors due to random sampling of subjects within group j. The expected value of ξ is the n x p null matrix. Different rows of $\xi.$ are statistically independent, but elements within rows are in general correlated and their covariance matrix is $\frac{1}{N_j}\Sigma$, where Σ is the covariance matrix of the p-variate vector observations.

Although X is of full column rank when the x_k are distinct (see Sec. III.B), A is in general not of full column rank. When A is of deficient column rank $r<m$, $|A'DA| = 0$ and, in consequence, the elements of β are not all estimable. Nevertheless, the normal equations arising from routine application of least-squares to (IV.4) are consistent, and the various mathematical methods of expressing their solution (such as placing independent restrictions on the rows of β) can be shown to be equivalent to decomposing the model matrix into

$$A = KL \quad , \qquad \text{(IV.5)}$$

where the n x r matrix K is a rank r column basis for A, and the r x m matrix L, also of rank r, contains the coefficients of certain selected linear parametric functions of the rows of β. (Bock, 1963; MSMBR, Sec. 5.1.) [When L is specified, K is obtained by $K = AL'(LL')^{-1}$.]

Substituting (IV.5) in (IV.4), and at the same time introducing the orthogonal reparameterization of the polynomials as in Section II, we may write

(IV. 4) as

$$Y. = K(L\beta S)P' + \xi.$$

$$= \underset{n\times r}{K}\ \underset{r\times s}{\Gamma}\ \underset{s\times p}{P'} + \xi. \qquad (IV.6)$$

Then the weighted least-squares (Gauss-Markov) estimator of Γ is (Lee, 1974)

$$\hat{\Gamma} = (K'DK)^{-1}\ K'DY.\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1} \quad . \qquad (IV.7)$$

The expected value of this estimator is Γ and its sampling variance-covariance matrix is given by the Kronecker product (see MSMBR, p. 212),

$$(K'DK)^{-1} \times (P'\Sigma^{-1}P)^{-1} \quad . \qquad (IV.8)$$

We notice, however, that (IV.7) contains the error variance-covariance matrix Σ and cannot be applied in general unless Σ is known. Fortunately, there are a number of straightforward methods of dealing with this problem:

1) Timm (1975, Sec. 5.16) has pointed out that if the degree of the polynomial model is set equal to p-1, then P is non-singular, $P^{-1} = P'$, and the matrix Σ^{-1} cancels out of (IV.7); i.e.,

$$\Sigma^{-1}P(P'\Sigma^{-1}P)^{-1} = \Sigma^{-1}PP'\Sigma P = P \ .$$

Thus, if p is not large and there is no advantage in using a less than p-1 degree trend polynomial, the Gauss-Markov estimator of the orthogonal coefficients is obtained in an unweighted analysis simply by transforming the vector observations by the p x p matrix of Fisher-Tshebycheff orthogonal polynomials P'.

2) If the structure of Σ is such that

$$\underset{p\times p}{\Lambda} = P'\Sigma P \qquad \text{IV.9}$$

is a diagonal matrix, the weight matrix cancels out of (IV. 7) for all values of $s \leq p$. This is true because, if P is an orthogonal matrix and (IV.9) is diagonal, the columns of P are the characteristic vectors of Σ and corresponding elements of Λ are the characteristic values associated with each. Thus, $\Sigma P = P\Lambda$, $\Sigma = P\Lambda P'$ and $\Sigma^{-1} = P'\Lambda^{-1}P$. Then if the $n \times s$ matrix P_1 contains the leading s columns of P, and the $s \times s$ diagonal matrix Λs contains the corresponding characteristic values, we have

$$\Sigma^{-1}P_1(P'\Sigma^{-1}P_1)^{-1} = P_1\Lambda_1^{-1}(\Lambda_1^{-1})^{-1} = P_1,$$

and the unweighted estimator is Gauss-Markov.

This result is of considerable practical interest because it can be shown under mixed-model assumptions (Bock, 1960, 1963; MSMBR, Sec. 7.1.2) that Σ belongs to a class of covariance structures diagonalized by a class of orthogonal transformations of which P is a member. The former is the class of so-called "reducible" covariance structures studied by Bargmann (1957; see also Huynh and Feldt, 1970).[3] Thus, unweighted multivariate analysis of repeated measures given in Bock (1963) and MSMBR (Chapter 7) is justified when applied under mixed-model assumptions, or more generally when $P'\Sigma P$ can be assumed diagonal, even when s is less than p.

This result also suggests that, in the presence of non-zero association in $P'\hat{\Sigma}P$, the transformed covariance matrix may be so greatly dominated by its diagonal elements that the unweighted estimates will differ but little from the weighted estimates.

The analysis in Example 2 is an instance in which this is the case.

3) Finally, if $P'\Sigma P$ is not diagonal, the most practical alternative would seem to be to forego an exact analysis and proceed by maximum likelihood estimation under large-sample assumptions. It can be shown (Khatri, 1966; Tubbs, Lewis & Duran, 1975) that, if a maximum likelihood estimator of Σ is available independent of Y. (e.g., the within-groups covariance matrix $\hat{\Sigma}$), then, for $|\Sigma| \neq 0$,

$$\hat{\hat{\Gamma}} = (K'DK)^{-1}K'DY.\hat{\Sigma}^{-1}P(P'\hat{\Sigma}^{-1}P)^{-1} \tag{IV.10}$$

is a consistent estimator of Γ with large-sample variance-covariance matrix given by the Kronecker product

$$(K'DK)^{-1} \times (P'\hat{\Sigma}^{-1}P)^{-1}. \tag{IV.11}$$

Associated with (IV.10) is a multivariate analysis of variance that plays the same role in longitudinal data as does the univariate analysis of variance of cross-sectional data in Section III B. For purposes of the multivariate analysis of variance, the columns may be orthogonalized from left to right with respect to D to obtain, say, K^*, where $K^{*\prime}DK^* = I_r$. Similarly, the columns of P_1 are orthogonalized from left to right with respect to $\hat{\Sigma}^{-1}$ to obtain, say P_1^*, where $P_1^{*\prime}\hat{\Sigma}^{-1}P_1^* = I_s$. Then, letting $M^* = K^{*\prime}D$ and $Q_1^* = \hat{\Sigma}^{-1}P_1^*$,

$$\underset{r\times s}{\hat{\hat{\Gamma}}^*} = M^*Y.Q_1^* = U = \begin{bmatrix} \underline{u}'_0 \\ \underline{u}'_1 \\ \vdots \\ \underline{u}'_{r-1} \end{bmatrix} \tag{IV.12}$$

is a maximum likelihood estimator of orthogonal parameters, Γ^*, with large-sample covariance matrix $I_r \times I_s$. The partition of the $s \times s$ matrix of sum of squares and cross-products (briefly "sums of products") for the multivariate analysis of variance may therefore be computed as shown in Table 7.

Insert Table 7 about here

The sums of products matrices are employed in tests of multivariate hypotheses of trend as follows:

Suppose it is desired to test the hypothesis that orthogonal polynomial coefficients greater than degree s_1-1 are null for some between-group effect represented by the sum of products $SSH = \sum_{k=c+1}^{c+n_h} SSB_k$ on n_h degrees of freedom. For this test, an error sum of products independent of SSH is extracted from the table. For example,

$$SSE^* = SSE + SSW \ ,$$

on $n_e = (N-r) \geq s-s_1$ degrees of freedom.

Then, for $s_2 = s-s_1$, the $s_2 \times s_2$ submatrices SSH_2 and SSE_2^* are extracted from the lower-right corners of SSH and SSE*, respectively, and the min (n_h, s_2) nonzero roots of the determinantal equation

$$|SSH_2 - \lambda SSE_2^*| = 0$$

are found and ordered from largest to smallest. From these roots, the following test statistics may be computed:

1) Roy's largest-root statistic, for example, in the form of the generalized F with arguments r and t (MSMBR, Sec. 3.4.7):

$$F_o = \frac{t}{r} \lambda_1 , \qquad \text{(IV.13)}$$

where $\lambda_1 = \max \lambda_\ell, \ \ell = 1,2,\ldots,\min(n_h, s_2)$

$r = n_h - s_2 + 1$

$t = n_e - s_2 + 1$.

Critical points for F_o may be read from the table for min(n_h, s_2) roots in Appendix A of MSMBR.

TABLE 7

Longitudinal Data: Multivariate Analysis of Variance of an s-1 Degree Polynomial Model for r Group Effects. The Sums of Squares and Cross-products are Computed from the Orthogonal Estimates $U = MY.Q'$

Source of Dispersion	Degrees of Freedom	Sums of Squares and Cross-products (sxs)
General mean	1	$SSM = \underline{u}_o\underline{u}_o'$
Between-group effects	1	$SSB_1 = \underline{u}_1\underline{u}_1'$
leading effects elim-	1	$SSB_2 = \underline{u}_2\underline{u}_2'$
inated and following	$\vdots$	
effects ignored	1	$SSB_{r-1} = \underline{u}_{r-1}\underline{u}'_{r-1}$
Between-group effects	r-1	SSB
Between-group residual	n-r	SSE = SSG - SSB - SSM
Group means	n	$SSG = Q[\sum_{j=1}^{n} N_j \underline{y}_{\cdot j} \underline{y}'_{\cdot j}]\, Q'$
Within Groups	N-n	$SSW = SST - SSG = I_s$
Total	$N = \sum_{j=1}^{n} N_j$	$SST = Q[\sum_{j=1}^{n} \sum_{i=1}^{N_j} \underline{y}_{ij} \underline{y}'_{ij}]\, Q'$

2) Hotelling trace statistic:

$$T_o^2 = n_e \Sigma_{\ell=1}^{\min(n_h, s_2)} \lambda_\ell \qquad \text{(IV.14)}$$

An F approximation for this statistic has been given by McKeon (1974):

$$F_M^{(K,D)} = \frac{mD}{s_2 n_h (D-2) n_e} T_o^2$$

where $m = n_e - s_2 - 1$

$K = s_2 n_h$

$D = 4 + \frac{K+2}{B-1}$

$B = \frac{(m+s_2)\ (m+n_h)}{(m-2)\ (m+1)}$

Exact .05 and .01 points for T_o^2 are available in Pillai (1960) and Pillai and Jayachandran (1970).

3) The likelihood ratio statistic:

$$\Lambda = \Pi_{\ell=1}^{\min(n_h, s_2)} \frac{1}{1 + \lambda_\ell} \qquad \text{(IV.15)}$$

Rao's F approximation (see MSMBR, Sec. 3.4.9) may be used to compute probability levels for Λ.

Note that, since the weighted analysis is based on an estimated Σ and assumes large-sample statistics, critical points of the distribution of $\chi^2/n_h s_2$ on $n_h s_2$ degrees of freedom (given, for example, by Hald, 1952) could be used in place of F_o, F_M, or F_R.

Both the weighted and the unweighted analysis include univariate F statistics for the separate terms of the polynomial. If the condition obtains that $P'\Sigma P$ is diagonal, these F statistics are statistically independent under multivariate normality. Thus, a union-intersection test, namely, that the hypothesis is rejected if the F_k is significant at the α_k level is a quite satisfactory multivariate test with joint significance level

$$\alpha^* = 1 - \Pi_{k=1}^{s_2} (1 - \alpha_k) \qquad \text{(IV.16)}$$

(c.f., Roy and Bargmann, 1958; J. Roy, 1958).

In the weighted case, where the terms are orthogonalized in the sample, (IV.16) applies in large samples. Because of the greater diagnostic and descriptive detail in the F's for the separate functions, the union-intersection test is in general more useful in repeated measures analysis than are statistics 1, 2 and 3 above. This is the test used in Example 2 in this section.

When the rank r of the model for the sampling or experimental effects and the rank s of the polynomial trend model have been chosen, possibly with the aid of the foregoing tests, estimates of the parameters in (IV.6) are obtained from (IV.12) as follows:

$$\hat{\hat{\Gamma}} = (S_r^{-1})' \, \hat{\hat{\Gamma}}^* T_s^{-1} \qquad \text{(IV.17)}$$

where S_r is the leading r rows and columns of the Cholesky factor of K'DK, and T_s is the s leading rows and columns of the Cholesky factor of $P'\hat{\Sigma}^{-1}P$ (see MSMBR, Sec. 2.7.2). These factors are given by the ORTHM subroutine of MATCAL (Brock and Repp, 1974) during the orthonormalization of K with respect to D and P with respect to $\hat{\Sigma}^{-1}$.

The fitted values of the group means may then be computed from

$$\hat{\hat{Y}}_{.} = K \underset{r \times e}{\hat{\hat{\Gamma}}} P' \; . \qquad \text{(IV.18)}$$

The variance-covariance matrix of the elements of (IV.18), rolled-out across rows, is

$$(K \times P)[(K'DK)^{-1} \times (P'\hat{\Sigma}^{-1}P)^{-1}](K \times P)' = K(K'DK)^{-1}K' \times P(P'\hat{\Sigma}^{-1}P)^{-1}P' \qquad \text{(IV.19)}$$

The 2 sigma tolerance interval on a new observation in group j at time point k is, therefore,

$$\hat{\hat{Y}}_{.j}^{(k)} \pm 2\sqrt{1 + [K]_j' \, (K'DK)^{-1}[K]_j \cdot \, [P]_k \, (P'\hat{\Sigma}^{-1}P)^{-1}[P]_k'} \quad , \qquad \text{(IV.20)}$$

where $[K]_j$ is the j-th row of K, written as a column, and $[P]_k$ is the k-th row of P, similarly written.

C. EXAMPLE 2: Longitudinal Comparison of Growth in Stature of Boys and Girls Aged 2 Through 8

Although sex differences in preadolescent stature are only weakly detected by the cross-sectional analysis in Example 1, they are clearly revealed in the present example when the inherently more powerful repeated measures analysis is brought to bear on longitudinal data. Table 8 summarizes measures of stature of boys and girls aged 2 through 8 from the Berkeley Guidance Study as reported by Tuddenham and Snyder (1954). Table 8, which includes standard deviations and product-moment correlations in addition to means and sample sizes, contains all of the information needed for a multivariate analysis of variance of mean trend in the two sex groups.

Insert Table 8 about here

The first step in examining these data is to test their conformity to the assumptions of the mixed model, unweighted, or weighted analysis. With the aid of the MULTIVARIANCE program, the common within-group covariance matrix is reconstructed from the standard deviations and correlations, and is transformed by the order 8 matrix of orthogonal polynomials in normalized form. (The MULTIVARIANCE program gives the user the option of such a transformation and generates the required matrix.) The result of this transformation is shown in Table 9.

Insert Table 9 about here

TABLE 8

Sample Statistics for Stature of 129 Complete Cases Aged 2 Through 8 from the Berkeley Guidance Study[a]

		Year Means (cm)						
		2	3	4	5	6	7	8
Group	N							
Boys	65	88.174	96.579	104.118	110.962	117.335	123.852	129.932
Girls	64	87.123	95.489	103.072	110.430	117.541	123.639	129.367
Within-group S.D.		3.1875	3.4445	3.8021	4.1688	4.5479	4.8241	5.1267
	2	1.0000						
	3	.8706	1.0000					
	4	.8243	.9352	1.0000				
Correlations	5	.8050	.9214	.9603	1.0000			
	6	.7821	.8904	.9187	.9701	1.0000		
	7	.7755	.8941	.9107	.9621	.9856	1.0000	
	8	.7583	.8790	.9047	.9474	.9723	.9889	1.0000

[a] These are the same cases studied by Thissen, et al. (1976)

TABLE 9
Orthogonal Polynomial Transformation of Statistics in Table 8

		Term						
		Constant	Linear	Quadratic	Cubic	Quartic	Quintic	Sextic
Group	N							
Boys	65	291.392	36.481	-1.9287	0.5175	-0.1188	-0.1366	-0.0626
Girls	64	289.771	37.324	-2.3025	-0.1531	-0.0995	-0.2169	0.0762
Within-Group	S.D.	10.5552	2.9507	1.2480	1.0299	0.7507	0.6770	0.5814
Correlations *	Const.	1.0000						
	Lin.	.6737	1.0000					
	Quad.	-.1934	-.2896	1.0000				
	Cubic	-.0031	-.0714	-.3716	1.0000			
	Quart.	-.0169	-.0732	-.1141	-.2413	1.0000		
	Quin.	.0658	.2170	.0790	-.3062	-.4563	1.0000	
	Sext.	-.0311	-.0090	.0208	-.0041	.1888	-.2531	1.0000

Characteristic roots of the within-group correlation matrix:

$$[\lambda_\ell] = [1.8899,\ 1.6498,\ 1.3464,\ 0.8845,\ 0.6907,\ 0.2911,\ 0.2475].$$

Inspecting the transformed correlations in Table 9, we see that the values in the first subdiagonal are large enough to suggest that the population matrix is not diagonal. We confirm this impression by calculating, from the characteristic roots of the correlation matrix shown in the footnote of Table 9, the likelihood ratio chi-square statistic for testing the hypothesis of no association in a p x p correlation matrix (Anderson, 1958, Chapter 9):

$$\chi^2_{p(p-1)/2} = -\left(N - n \frac{2p+5}{6}\right) \sum_{\ell=1}^{p} \ln\lambda_\ell \qquad \text{(IV.21)}$$

$$= 209.10$$

On the null hypothesis and multivariate normal distribution of the within-group residuals, this statistic is distributed in large samples as a central chi-square variate on p(p-1)/2 degrees of freedom. In this instance, the number of degrees of freedom is 21 and the value of the chi-square clearly contradicts the hypothesis that the correlation matrix (and thus the covariance matrix) is diagonal. We therefore conclude that a weighted analysis is necessary.

Had we accepted the hypothesis that the population covariance matrix is diagonal, we would have tested the variances of the linear through sextic terms for homogeneity. [The Hartley variance-range test is suitable for this purpose (Pearson and Hartley, 1966, p. 202).] If these variances had appeared homogeneous, we could have then pooled them to obtain for the R ratios a denominator with 6 x 127 = 762 degrees of freedom. Because of the resulting increase in power of the F tests, this "mixed-model" form of analysis is obviously the preferred approach to repeated measures data when it is justified (see Bock, 1963; Huynh & Feldt, 1970).

If the population covariance matrix is assumed diagonal but the variances in multiple-degree of freedom subspaces of the within-subject variation are not assumed homogeneous, the unweighted repeated measures analysis is indicated. This type of analysis is simply a multivariate analysis of variance of the orthogonal polynomial transform of the original data. In the MULTIVARIANCE program, it is performed by transforming the summary statistics before entering the estimation of tests-of-hypothesis phases of the program. The estimation phase computes the basis matrix K in (IV.6) and estimates effects associated with the experimental or sampling structure of the data. In the present study, the sampling structure consists simply of the classification of the subjects as male and female. Since there are only two sample groups, the K matrix generated by the program is

$$K_2 = \begin{bmatrix} 1 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{bmatrix} .$$

The first column of K_2 corresponds to the one degree of freedom for the general mean, ignoring the sex classification, and the second column corresponds to the one degree of freedom between groups. Associated with each of these degrees of freedom is an F statistic for each of the terms in the polynomial; these statistics are independent if the transformed covariance matrix $P'\Sigma P$ is diagonal.

When the transformed covariance matrix is not diagonal and the weighted analysis is required, an additional step must be interposed between the calculation of the summary statistics and the multivariate analysis of variance: the matrix of orthogonal polynomials must be orthogonalized again with respect to the inverse sample covariance matrix. The MATCAL subroutine ORTHM performs this operation (Bock & Repp, 1974). For the present data, this inverse is shown in Table 10. The polynomials orthogonalized with respect to this matrix are shown in Table 11. The (upper triangular) matrix of the transformation of the order 7 orthogonal polynomials (i.e., the T_s^{-1} matrix of formula IV.17) is

----

Insert Tables 10 & 11 about here

TABLE 10

Inverse of the Sample Within-group Covariance Matrix

	1	2	3	4	5	6	7
1	.4154				(Symmetric)		
2	-.3141	1.0020					
3	-.0570	-.4520	1.2923				
4	.0564	-.0905	-.0244	2.1499			
5	-.0989	.2362	.0192	-.8347	2.2599		
6	.0127	-.4584	.6559	-.6421	-.6296	4.2315	
7	.0579	.1522	-.4108	.3753	.1045	-2.2124	1.8759

TABLE 11

Orthogonal Polynomials for Weighted Trend Analysis

			Terms				
Age	Constant	Linear	Quadratic	Cubic	Quartic	Quintic	Sextic
2	2.9204	1.0426	.3385	-.5765	.2984	-.0876	-.0211
3	"	1.5517	-.4531	.4506	-.5169	.3225	-.3853
4	"	2.0609	-1.0041	.6411	-.0364	-.4368	.2625
5	"	2.5700	-1.3146	.3519	.4358	.0383	-.4495
6	"	3.0791	-1.3844	-.0601	.3468	.7924	.0688
7	"	3.5883	-1.2137	-.2379	-.1057	.5130	-.2525
8	"	4.0974	-.8024	.1754	.0268	.5308	-.0921

given in Table 12. Tables 10 and 11 contain the matrices required for the computation

----

Insert Table 12 about here

----

of the transformation matrix for the weighted analysis as given by $Q^* = \hat{\Sigma}^{-1}P^*$. Since the MULTIVARIANCE program has a provision for any arbitrary linear transformation of the sample data, it is a simple matter to perform the weighted analysis with the transformation matrix Q^*.

For purposes of comparison, the results of both the weighted and unweighted repeated measures trend analysis are shown in Table 13. Despite the apparent smooth progression of sample means shown in Table 8, the differences between the sex groups is not a simple function of age. Both the weighted and unweighted analysis show a significant degree 5 trend component in the differences between the means of the sex groups. To represent completely systematic differences in average stature of boys and girls in this age range therefore requires a rank 2 model for sample classes (r=2) and a rank 6 model for trend (s=6).

Insert Table 13 about here

If the subjects regarded as a sample from a single population and the sex groups are combined, the curve of mean growth is considerably simpler. Both analyses show at most a marginally significant cubic component. Ignoring the sex classification thus leads to a rank 1 (r=1) model for the sample and a rank 4 (s=4) model for trend.

The weighted and unweighted estimates of the orthogonal polynomial coefficients for these models are shown in Table 14. Note that when the quintic

TABLE 12

Coefficient Transformation From Unweighted to Weighted Orthogonal Polynomials

		Unweighted						
		1	2	3	4	5	6	7
	1	7.7266	6.7996	-2.2050	0.2813	0.1697	0.6322	-.3285
	2		2.6941	-1.0062	0.0336	0.0739	0.6549	-.0267
	3			1.1025	-.5626	-.1144	.1086	.0260
Weighted	4				.8742	-.4352	-.3271	-.0042
	5					.6656	-.3170	.1417
	6						.6550	-.1714
	7	(Triangular)						.5814

TABLE 13

Test Statistic for Longitudinal Trend Analysis of Growth in Stature of Children Aged 2 Through 8 in the Berkeley Guidance Study (Within-groups df=127)

Source of Variation	d.f.	Weighted		Unweighted	
		F	p	F	p
General Mean	1				
Constant		84,805	<.0001	97,770	<.0001
Linear		21,770	<.0001	20,172	<.0001
Quadratic		443.0	<.0001	370.2	<.0001
Cubic		3.576	.061	4.154	.043
Quartic		2.420	.122	2.729	.101
Quintic		0.4969	.482	.4223	.517
Sextic		.0150	.903	.0150	.903
Between Sexes	1				
Constant		4.210	.042	.762	.385
Linear		.0867	.769	2.631	.107
Quadratic		11.040	.001	2.893	.091
Cubic		5.388	.022	13.672	.004
Quartic		2.264	.135	.0213	.884
Quintic		11.692	.001	8.792	.004
Sextic		1.836	.178	1.836	.178

Insert Table 14 about here

----

polynomial is assumed, the weighted and unweighted solutions give essentially the same result. This is a consequence of the very small correlations between the sixth degree term and the remaining terms of the polynomial (Table 9). With respect to the last term, the transformed error covariance matrix is nearly diagonal, and the weighted and unweighted analysis become nearly identical at the fifth degree term.

When the cubic polynomial is assumed, on the other hand, the effects of the weights are more evident, especially in the sex effect and sex x age interactions, which are sensitive to the greater weight assigned at younger ages where the within-sex variation is smaller. Despite the significant failure of fit of the degree-3 model, the fitted means reproduce the observed means almost as well as the much less parsimonious degree-5 model. This is apparent in Table 5, where the observed means in Table 8 are reproduced along with means computed by (IV.18) from the weighted estimates in Table 14. The figures in Table 15 demonstrate the efficacy of low-degree polynomial models for growth when a limited age span is examined.

Insert Table 15 about here

V. SUMMARY

A methodological problem widely encountered in the study of secular trend, growth and development, or experimental manipulation of behavior is that of detecting and describing systematic change over time. With certain restrictions on the design of the study and the method of measurement, this problem has a ready solution in analysis of variance and its multivariate extensions. To be

TABLE 14

Longitudinal Trend Analysis: Estimated Orthogonal Polynomial Coefficients

Effect	Weighted		Unweighted
	Degree 5	Degree 3	
General Mean			
Constant	290.5870	290.5694	290.5816
Linear	36.9030	36.8709	36.9026
Quadratic	-2.1160	-2.1386	-2.1156
Cubic	0.1823	.1442	.1822
Quartic	-0.1109		-.1091
Quintic	0.0422		.0401
Sex (B-G)	1.5436	1.9692	1.6220
Age x Sex			
Linear	-0.8492	-0.4353	-0.8429
Quadratic	0.3800	.4151	.3738
Cubic	0.6696	.3573	.6706
Quartic	0.0145		-.0193
Quintic	-0.3944		-.3535

TABLE 15

Observed and Fitted Age-group Means for the Berkeley Guidance Study Data

	Mean Stature (cm.)					
	Observed		Fitted[a]			
			Degree 5		Degree 3	
	Boys	Girls	Boys	Girls	Boys	Girls
Age (yrs.)						
2	88.17	87.12	88.17	87.13	88.23	87.16
3	96.58	95.49	96.54	95.54	96.48	95.42
4	104.12	103.07	104.15	103.04	104.03	103.20
5	110.96	110.43	110.91	110.49	111.04	110.48
6	117.34	117.54	117.34	117.53	117.62	117.24
7	123.85	123.64	123.83	123.67	123.92	123.48
8	129.93	129.37	129.92	129.38	130.06	129.18

[a] From the weighted estimates

amenable to this technique, the observations should be time-structured (i.e., limited to a moderate number of pre-assigned time points, preferably equally spaced), and the measurement of the trait or response in question should be made on a scale with commensurate units throughout the relevant range. The analysis is further facilitated if, in cross-sectional data, the observations are replicated at each time point, or, in longitudinal data, all subjects are observed at precisely the same or comparable time points.

When the data are cross-sectional (i.e., independent samples of subjects are drawn contemporaneously at the several time points), the shape of the curve describing time-dependency of the population mean, and differences in the shape between populations, can be investigated in a polynomial trend analysis. A single-degree-of-freedom univariate analysis of variance of successive orthogonal polynomial components of trend and trend differences provides a convenient, exact, unbiased minimum-variance method of performing this analysis. Although the calculations for this analysis are most straightforward when the time-points are equally spaced and the equal numbers of subjects are sampled at each time point, the statistical theory and computer methods for unequal spacing and unbalanced sampling are fully worked out and available if needed.

When the data are longitudinal (i.e., each subject is measured at each time point), trend analysis of the population time-point means, or differences in trend between populations, can be carried out by multivariate repeated measures analysis or, in favorable cases, by mixed-model univariate analysis of variance. If the variance-covariance structure of the sampling errors is transformed to no-association (i.e., uncorrelated) by a suitable orthogonal matrix independent of the data (e.g., a matrix of Fisher-Tchebycheff orthogonal polynomials), an exact analysis of trend is provided by a multivariate analysis of variance in which the orthogonal components of trend appear as variates but are tested in a manner analogous to, but in general more powerful than, the single-degree-of freedom tests in the univariate analysis of variance of cross-sectional data. If the transformed errors are uncorrelated and the error trend components exclusive of the constant component are homogeneous in variance, a pooled estimate of the error components may be used and the multivariate repeated measures analysis specializes to a still more powerful single-degree-of-freedom mixed model univariate analysis of variance.

If the error covariance structure cannot be reduced to no association by a suitable orthogonal matrix independent of the data, a consistent, efficient, large-sample, weighted method of repeated measures analysis based on the Pottoff-Roy formulation may be available. It is shown here that this analysis is conveniently implemented by orthogonalizing the Fisher-Tshebycheff orthogonal polynomials with respect to the inverse of the sample within-group covariance matrix. The statistical tests of conventional multivariate analysis of variance of trend components computed with this re-orthogonalized matrix weighted by the inverse sample covariance matrix may then be interpreted in a large-sample sense.

Examples of these procedures applied to measures of stature and computed with the MULTIVARIANCE program are presented in the text.

Footnotes

[1] This section is based on Section 5.2.5 of Bock, R. D. *Multivariate statistical methods in behavioral research*. McGraw-Hill, 1975 (hereafter referred to as MSMBR).

[2] Various designs for semi-longitudinal studies have been proposed (Schaie, 1965). Their analysis is beyond the scope of this paper.

[3] The hypothesis that the population transformed error matrix is diagonal may be tested by a likelihood ratio test of no-association in the transformed sample matrix $P'\hat{\Sigma}P$. (See Anderson, 1958, Chapter 9.)

THE ANALYSIS OF CATEGORICAL DATA IN LONGITUDINAL STUDIES OF BEHAVIORAL DEVELOPMENT

by

J. Richard Landis
Department of Biostatistics
School of Public Health
University of Michigan
Ann Arbor, Michigan 48109

and

Gary G. Koch
Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, North Carolina 27514

Table of Contents

THE ANALYSIS OF CATEGORICAL DATA IN LONGITUDINAL STUDIES OF BEHAVIORAL DEVELOPMENT

J. Richard Landis
University of Michigan

Gary G. Koch
University of North Carolina

ABSTRACT
I. INTRODUCTION
II. LONGITUDINAL DATA EXAMPLES
III. METHODOLOGY
IV. ANALYSIS OF LONGITUDINAL DATA EXAMPLES
V. DISCUSSION
VI. SUMMARY

ACKNOWLEDGMENTS

The authors would like to thank Professors John R. Nesselroade and Paul B. Baltes for helpful comments and suggestions during preliminary discussions of this topic. In addition, the authors are grateful to Ms. Karen A. Stanecki and Ms. Roberta K. Selin for their competent assistance in the data processing associated with the examples, and to Ms. Linda L. Blakley for her cheerful and conscientious typing of the manuscript.

ABSTRACT

This paper is concerned with the analysis of multivariate categorical data which are obtained from longitudinal studies of human growth and development. An expository discussion of pertinent hypotheses for such situations is provided within the context of two methodologically illustrative examples, and appropriate test statistics are developed through the application of weighted least squares. These procedures are illustrated with extensive analyses of each of the data sets.

I. INTRODUCTION

Many longitudinal investigations dealing with behavioral and/or educational development are concerned with intraindividual change for variables which are measured in terms of discrete categories (based on nominal or ordinal as opposed to interval scales). Examples of such categorical (qualitative) variables include measures of

(i) child competencies in task performance,
(ii) verbalization patterns,
(iii) coping ability in stress situations,
(iv) self-concept,
(v) parent-child interaction,
(vi) developmental stages.

Thus, the basic research design for such studies involves the classification of each subject with respect to each categorical variable (which is called an attribute) at each of several successive time points. Moreover, the subjects may also be classified into a set of sub-populations on the basis of other categorical variables such as

(vii) demographic characteristics,
(viii) health status with respect to the presence or absence of certain traits, e.g., dyslexia,
(ix) program status with respect to certain specialized types of training.

Within the context of this framework, the data resulting from longitudinal studies can be conceptually arrayed in (potentially very large) multidimensional contingency tables for which the corresponding dimensions are the classifications according to each attribute like (i)-(vi) at each time point, together with the respective sub-population variables like (vii)-(ix). For this reason, the various questions of interest for data analysis may be regarded as equivalent to statistical models for the cell probabilities in such contingency tables. Thus, in principle, data analysis can be undertaken in terms of various computational algorithms for contingency table model fitting. As will be demonstrated in Section III, the weighted least squares methods discussed by Grizzle *et al*. (1969), Koch *et al*. (1977), and Landis *et al*. (1976) can be used to deal with questions pertaining to

(a) the nature and extent of intraindividual change for the respective attributes over time, both separately and simultaneously,

(b) the nature and extent of interindividual differences among two or more sub-populations with respect to intraindividual change over time for the respective attributes,

(c) the nature and extent of the variation over time of the relationship among two or more attributes as reflected by certain measures of association.

In this regard, the basic approach is in the same spirit as multivariate analysis of variance (profile analysis and/or growth curve analysis) with respect to analogous intervally-scaled quantitative data situations. For this reason, its application to longitudinal data is the primary purpose of this paper. The types of examples where weighted least squares methods are potentially most useful are briefly described in Section II. Specific formulations of the various hypotheses of interest are then presented in

Section III and their evaluation is illustrated in Section IV with respect to the examples in Section II. The paper concludes with a brief discussion of certain special computational considerations for dealing with very large contingency tables, together with other potential difficulties which may arise in the analysis of longitudinal data.

Alternatively, in many investigations, the questions of interest pertain to the identification of the underlying structure of a large number of attributes in terms of a small number of implicit dimensions, and hence are analogous to those for which factor analysis is used with quantitative data. This topic is outside the scope of the present paper. However, it is discussed to some extent by Goodman (1974) in the context of latent structure analysis. Similarly, the analysis of transition patterns over time for one or more attributes in terms of stochastic process models is also outside the scope of this paper. Further information with respect to this general subject area is given in Bishop, Fienberg, and Holland (1975, Chapter 7).

II. LONGITUDINAL DATA EXAMPLES

In this section, we present two examples which may be regarded as methodologically illustrative data sets from longitudinal studies in human growth and development. In particular, a simple one-population study involving two attributes measured at two time points is considered in Section II A to indicate the full range of hypotheses which can be tested. Moreover, in Section II B a two-population study involving one attribute measured at three points in time is used to illustrate the types of hypotheses associated with comparing growth curves among several groups. Although these examples involve hypothetical data and are much smaller in scope than those usually encountered in research situations, the extension to more complex designs is straightforward as developed in Section III.

A. A Single Population Example

The following hypothetical example arose from a longitudinal study in which two developmental attributes labelled A1 and A2 were measured at two time points labelled T1 and T2 for an age cohort of 354 children. In this regard, each subject was classified as absent (1) or present (2) for each of the attributes at each of the specified time points. The frequency data corresponding to each of the 16 possible response profiles is shown in Table 1.

Insert Table 1 About Here

The statistical issues concerning intraindividual change can be summarized within the framework of the following basic questions.

(1) Are there any differences between the occurrence rates of the two attributes at each of the time points?

(2) Are there any differences between the two time points with respect to the set of individual occurrence rates of the two attributes?

(3) Is there any attribute x time interaction in the occurrence rates of the two attributes?

Table 1

ATTRIBUTE DATA FOR LONGITUDINAL STUDY

		Response Profile for A1 and A2 at T1 and T2																TOTAL
T1	A1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	
	A2	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	
T2	A1	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	
	A2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	
FREQUENCY		57	36	18	69	0	0	0	33	0	3	0	15	0	0	0	123	354

(4) Are there any differences between the two time points with respect to the overall joint distribution of the two attributes?

(5) Is there any difference between the two time points with respect to a selected measure of association or agreement between the two attributes?

The first three questions which involve the occurrence rates of the two attributes are essentially similar to the hypotheses of interest in repeated measurement (or mixed model) experiments as discussed in further detail in Koch and Reinfurt (1971), Koch _et al_. (1977) and Landis and Koch (1977a). More specifically, question (1) addresses differences among attributes, question (2) involves the issue of time point differences, and question (3) is concerned with the attribute x time interaction as measured by the individual occurrence rates of the attributes. Thus, the first-order (univariate) marginal distributions of response for each of the attributes within each time point contain the relevant information for dealing with these questions. In contrast to overall average differences among the occurrence rates, questions (4)-(5) address the relationship between the attributes on specific subjects across the time periods. As a result, these questions involve measures of association or agreement between the attributes such as those discussed in Bishop, Fienberg, and Holland (1975) and Landis and Koch (1975a, 1975b). Hence, certain functions of the diagonal cells of various subtables are used to provide information for dealing with these questions.

B. A Two-Population Example

The following hypothetical example arose from a longitudinal study to compare boys and girls from a selected age cohort with respect to their ability to perform a particular behavioral task. In this regard, each subject was graded as success (S) or failure (F) at the end of 1 year, 2 years, and 4 years of follow-up. These resulting data are shown in Table 2.

Insert Table 2 About Here

Accordingly, the statistical issues concerning these differences in growth patterns can be summarized within the framework of the following basic questions:

(1) Are there any differences between the boys and the girls with respect to the behavioral task success rates at the three time points?

(2) Are there any differences among the three time points with respect to the behavioral task success rates across the two groups of children?

(3) Is there any sex group x time interaction with respect to the behavioral task success rates?

These three questions involving the success rates are directly analogous to the hypotheses of "no whole-plot effects," "no split-plot effects," and "no whole-plot x split-plot interaction" in standard split-plot experiments as described in Anderson and Bancroft (1952), Federer (1955), or Steel and Torrie (1960). In particular, since time is the split-plot factor, these resulting success rates give rise to growth profiles for each sex group. In this context, question (1) addresses group differ-

Table 2

LONGITUDINAL DATA FOR BEHAVIORAL TASK

	Response Profile at year 1 vs year 2 vs year 4								
Sex	SSS	SSF	SFS	SFF	FSS	FSF	FFS	FFF	TOTAL
Boys	0	0	5	0	21	14	51	71	162
Girls	16	5	12	5	71	12	74	14	209

ences, question (2) involves the issue of time differences, and question (3) is concerned with the hypothesis of parallelism among corresponding segments of the growth profiles. Consequently, the joint set of first-order (univariate) marginal distributions for each of the time points within each sex group contain the relevant information for dealing with these questions.

III. METHODOLOGY

This section is concerned with a general methodology for answering the types of questions outlined in Section II in terms of specific hypotheses. Because the measurement scales of the response variables (hereafter referred to as attributes) are categorical, the conceptual formulation of such hypotheses must be undertaken in terms of an underlying (s x r) contingency table, where s is the number of sub-populations and r is the number of possible multivariate response profiles. Test statistics for such hypotheses and the estimators for parameters of underlying linear regression models are obtained through weighted least squares computations by methods originally described in Grizzle, Starmer, and Koch (1969) (hereafter referenced as GSK) as reviewed in the Appendix. Consequently, this methodology represents a categorical data analogue to more well-known counterparts for quantitative data like multivariate analysis of variance as described by Cole and Grizzle (1966) and Morrison (1967) in the parametric case and multivariate rank analysis as described by Koch (1969, 1970) in the non-parametric case.

For longitudinal studies, each subject is measured on the same set of d attributes at each of t time points. In accordance with the general framework in the Appendix, let i = 1,2,...,s index a set of sub-populations from which random samples have been selected. Let $m = 1,2,\ldots,d$ index a set of d characteristics or attributes corresponding to behavioral tasks which are measured at each time point using an L_m-point scale. Then let the $r = (L_1 L_2 \ldots L_d)^t$ response profiles be indexed by a vector subscript $\underset{\sim}{j} = (\underset{\sim}{j}_1, \underset{\sim}{j}_2, \ldots, \underset{\sim}{j}_t)$ with $\underset{\sim}{j}_g = (j_{g1}, j_{g2}, \ldots, j_{gd})$, where $j_{gm} = 1,2,\ldots,L_m$ for $m = 1,2,\ldots,d$ and $g = 1,2,\ldots,t$. Furthermore, let $\pi_{i\underset{\sim}{j}} = \pi_{i\underset{\sim}{j}_1,\underset{\sim}{j}_2,\ldots,\underset{\sim}{j}_t}$ represent the joint probability of response profile $\underset{\sim}{j}$ for randomly selected subjects from the i-th sub-population. Then the first-order marginal probability

$$\phi_{igmk} = \sum_{\underset{\sim}{j} \text{ with } j_{gm}} \cdots \sum_{= k} \pi_{i\underset{\sim}{j}_1,j_2,\ldots,\underset{\sim}{j}_t} \quad \text{for} \quad \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ m = 1,2,\ldots,d \\ k = 1,2,\ldots,L_m \end{array} \qquad \text{(III.1)}$$

represents the probability of the k-th response category of the m-th attribute at the g-th time point in the i-th sub-population.

In addition, it should be noted here that this formulation for longitudinal studies presumes the following conditions:

(i) there is no assumed structure on the attributes, so that all possible attribute combinations or response profiles are observable, i.e., $\pi_{i\underset{\sim}{j}} > 0$ for all $\underset{\sim}{j}$ and for $i=1,2,\ldots,s$;

(ii) every subject entering the study is followed until completion of the study, i.e., there are no lost-to-follow up cases;

(iii) every subject is measured on each attribute at each time point, i.e., there is no incomplete data.

Further discussion of these potential difficulties is given in Section V in the form of concluding remarks.

A. Hypotheses Involving Marginal Distributions

Hypotheses directed at questions pertaining to average differences among sub-populations, attributes, and time points involve the first-order marginal distributions of the response profiles and can be expressed in terms of constraints on the corresponding probabilities $\{\phi_{igmk}\}$. More specifically, the hypotheses associated with questions (1) - (3) of each of the examples in Section II can be formulated within the scope of one of the following statements:

(1) If there are no differences among the marginal distributions of the respective attributes at each time point for the s sub-populations, then the $\{\phi_{igmk}\}$ satisfy the hypothesis

$$H_{SM}: \quad \phi_{1gmk} = \phi_{2gmk} = \cdots \phi_{sgmk} \quad \text{for } \begin{array}{l} g = 1,2,\ldots,t \\ m = 1,2,\ldots,d \\ k = 1,2,\ldots,L_m \end{array}, \qquad (\text{III}.2)$$

where SM denotes sub-populations means;

(2) If there are no differences among the marginal distributions of the respective attributes over the t time points within each of the sub-populations, then the $\{\phi_{igmk}\}$ satisfy the hypothesis of composite first-order marginal homogeneity (symmetry)

$$H_{TMH}: \quad \phi_{i1mk} = \phi_{i2mk} = \cdots = \phi_{itmk} \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ m = 1,2,\ldots,d \\ k = 1,2,\ldots,L_m \end{array}, \qquad (\text{III}.3)$$

where TMH denotes time marginal homogeneity;

(3) If there is no time x subgroup interaction (with respect to the marginal distributions of the respective attributes at the t time points), then the $\{\phi_{igmk}\}$ may be written in terms of an additive model

$$H_{ST}: \quad \phi_{igmk} = \mu_{mk} + \xi_{i*mk} + \tau_{*gmk} \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ m = 1,2,\ldots,d \\ k = 1,2,\ldots,L_m \end{array}, \qquad (\text{III}.4)$$

where ST denotes subgroup x time, and where for the m-th attribute, μ_{mk} is an overall mean associated with the k-th response category, ξ_{i*mk} is an effect due to the i-th sub-population, and τ_{*gmk} is an effect due to the g-th time point, and where it is understood that the $\{\mu_{mk}\}$, $\{\xi_{i*mk}\}$, and $\{\tau_{*gmk}\}$ satisfy the usual analysis of variance constraints.

Moreover, if the d attributes are all measured on the same L-point scale, it follows that $L_m = L$ for $m = 1,2,\ldots,d$. For example, each attribute may be classified as present or absent as proposed in the example in Section 2.1, or each attribute may represent a different scheme of classifying development under the assumption that there exists an identical number of steps or stages as discussed in Wohlwill (1973). In such situations, several additional hypotheses of this type may become of interest.

(4) If there are no differences among the marginal distributions of the attributes at each of the time points within each of the sub-populations, then the $\{\phi_{igmk}\}$ satisfy the hypothesis of marginal homogeneity (symmetry) among the attributes

$$H_{AMH}: \quad \phi_{ig1k} = \phi_{ig2k} = \cdots \phi_{igmk} \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ k = 1,2,\ldots,L \end{array} \quad , \qquad \text{(III.5)}$$

where AMH denotes attribute marginal homogeneity;

(5) If there is no interaction between the marginal distributions of the attributes and time within each sub-population, then the $\{\phi_{igmk}\}$ may be written in terms of an additive model

$$H_{AT}: \quad \phi_{igmk} = \mu_{ik} + \xi_{ig*k} + \tau_{i*mk} \qquad \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ m = 1,2,\ldots,d \\ k = 1,2,\ldots,L \end{array} , \qquad \text{(III.6)}$$

where AT denotes attribute x time, and where for the i-th sub-population, μ_{ik} is an overall mean associated with the k-th response category, ξ_{ig*k} is an effect due to the g-th time point, and τ_{i*mk} is an effect due to the m-th attribute, and where it is understood that the $\{\mu_{ik}\}$, $\{\xi_{ig*k}\}$, and $\{\tau_{i*mk}\}$ satisfy the usual analysis of variance constraints.

All of these considerations can be extended somewhat further if the response categories $k = 1,2,\ldots,L_m$ for $m = 1,2,\ldots,d$ are ordinally scaled with progressively larger intensities. In this situation, the effects of the respective sub-populations, attributes, and time points can be compared in terms of summary indexes

$$\eta_{igm} = \sum_{k=1}^{L_m} a_{mk}\phi_{igmk} \quad \text{for } \begin{matrix} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ m = 1,2,\ldots,d \end{matrix} . \qquad \text{(III.7)}$$

Here η_{igm} can be regarded as a mean score for the m-th attribute at the g-th time period in the i-th sub-population with respect to an underlying numerical scaling $a_{m1}, a_{m2}, \ldots, a_{mL_m}$ of the L_m categories. In this context, the $\{\eta_{igm}\}$ are equivalent to mean scores derived from strictly quantitatively scaled response categories as discussed in Bhapkar (1966). Thus, the hypotheses in (III.2 - III.6) can also be expressed in terms of constraints on the $\{\eta_{igm}\}$ in (III.7). Expressions of these hypotheses in terms of the $\{\eta_{igm}\}$ are discussed in more detail in Koch _et al_. (1977) and are illustrated in Landis (1975).

B. Hypotheses Involving Measures of Association

Whereas the hypotheses in Section III A were addressed at comparisons among sub-populations, time points, and attributes within the context of first-order marginal distributions, the hypotheses in this section are directed at relationships among the attributes at a given time point, and the extent to which those relationships change across time. These hypotheses can be formulated in terms of comparisons among full joint distributions or second and higher-order joint marginal distributions across time periods or in terms of measures of association such as the log cross-product ratio for nominal data as discussed in Bhapkar and Koch (1968a, 1968b) or the Goodman-Kruskal rank correlation coefficient for ordinal data as discussed in Forthofer and Koch (1973).

In general, these hypotheses can be expressed as a set of constraint equations on the joint probabilities of specified response profiles. For purposes of simplicity, we will focus on the joint distributions of two selected attributes (relabelled as 1 and 2) at each of the t time points for each of the s sub-populations. Consequently, the joint probability of the k_1-th category on the m_1-th attribute and the k_2-th category on the m_2-th attribute at the g-th time point in the i-th sub-population can be written as

$$\Psi_{igk_1k_2} = \sum_{\underset{\sim}{j} \text{ with } j_{gm_1} = k_1} \cdots \sum_{\text{and } j_{gm_2} = k_2} \pi_{i\underset{\sim}{j}_1,\underset{\sim}{j}_2,\ldots,\underset{\sim}{j}_t} \; . \qquad \text{(III.8)}$$

Using this notation, the log cross-product measures of association between the two attributes can then be expressed as

$$\Delta_{igk_1k_2} = \log_e \left\{ \frac{\Psi_{igk_1k_2}\,\Psi_{igL_1L_2}}{\Psi_{igL_1k_2}\,\Psi_{igk_1L_2}} \right\} \quad \text{for} \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ k_1 = 1,2,\ldots,L_1-1 \\ k_2 = 1,2,\ldots,L_2-1 \end{array} . \qquad \text{(III.9)}$$

In particular, for two dichotomous attributes (L_1=L_2=2), the measures of association $\{\Delta_{igk_1k_2}\}$ in (3.9) reduce to the familiar log cross-product ratio for a 2 x 2 table

$$\Delta_{ig} = \log_e \left\{ \frac{\Psi_{ig11}\,\Psi_{ig22}}{\Psi_{ig21}\,\Psi_{ig12}} \right\} \quad \text{for} \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \end{array} . \qquad \text{(III.10)}$$

Otherwise, an alternative measure of association for 2 x 2 tables due to Yule can be formulated as

$$Q_{ig} = \frac{\Psi_{ig11}\Psi_{ig22} - \Psi_{ig21}\Psi_{ig12}}{\Psi_{ig11}\Psi_{ig22} + \Psi_{ig21}\Psi_{ig12}} \quad \text{for} \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \end{array} . \qquad \text{(III.11)}$$

(For a more complete discussion of measures of association, see Bishop, Fienberg, and Holland, 1975, Chapter 11.)

Hypotheses concerning comparisons among full joint distributions of the attributes can now be expressed in terms of constraints on the joint probabilities $\{\Psi_{igk_1k_2}\}$. More specifically, hypotheses associated with questions such as (4) in Section II A can be formulated within the scope of one of the following statements:

(6) If there are no differences among the joint distributions of the two attributes among the sub-populations, then the $\{\Psi_{igk_1k_2}\}$ satisfy the hypothesis

$$H_{SJD}: \Psi_{1gk_1k_2} = \Psi_{2gk_1k_2} = \ldots \Psi_{sgk_1k_2} \quad \text{for } \begin{array}{l} g = 1,2,\ldots,t \\ k_1 = 1,2,\ldots,L_1 \\ k_2 = 1,2,\ldots,L_2 \end{array} \qquad \text{(III.12)}$$

where SJD denotes sub-population joint distributions;

(7) If there are no differences among the joint distributions of the two attributes among the time points, then the $\{\Psi_{igk_1k_2}\}$ satisfy the hypothesis

$$H_{TJD}: \Psi_{i1k_1k_2} = \Psi_{i2k_1k_2} = \ldots = \Psi_{itk_1k_2} \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ k_1 = 1,2,\ldots,L_1 \\ k_2 = 1,2,\ldots,L_2 \end{array} \qquad \text{(III.13)}$$

where TJD denotes time joint distributions.

Additional hypotheses involving the joint distribution probabilities in (III.8), such as an additive model implying no interaction between sub-populations and time periods directly analogous to (III.4), could also be considered here. Moreover, hypotheses of "no interaction" among higher-order joint distributions involving more than two attributes simultaneously can be developed as direct extensions of these results, although the notation for corresponding expressions becomes more cumbersome.

Similar considerations also apply to hypotheses of "no interaction" for the joint distribution over time of each separate attribute and/or simultaneous sets of attributes. Finally, a log-linear model can be fitted to the joint distribution of the attributes at each time point under appropriate hypotheses of "no interaction" as discussed in Koch et al. (1976). This approach then permits hypothesis testing for relationships across the time points in terms of the resulting log-linear model parameters.

Alternatively, several hypotheses associated with questions such as (5) in Section (II.A) involving the measures of association in (III.9) can be formulated as follows:

(8) If the two selected attributes are independent of each other at each time point within each sub-population, then the $\{\Delta_{igk_1k_2}\}$ satisfy the hypothesis

$$H_{PI}: \ \Delta_{igk_1k_2} = 0 \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ k_1 = 1,2,\ldots,L_1-1 \\ k_2 = 1,2,\ldots,L_2-1 \end{array} , \quad (III.14)$$

where PI denotes pairwise independence;

(9) If the relationship between the two attributes as measured by the log cross-product ratio is the same across the time points for each sub-population, then the $\{\Delta_{igk_1k_2}\}$ satisfy the hypothesis.

$$H_{TA}: \ \Delta_{i1k_1k_2} = \Delta_{i2k_1k_2} = \ldots = \Delta_{itk_1k_2} \quad \text{for } \begin{array}{l} i = 1,2,\ldots,s \\ k_1 = 1,2,\ldots,L_1-1 \\ k_2 = 1,2,\ldots,L_2-1 \end{array} , \quad (III.15)$$

where TA denotes time association;

(10) If the relationship between the two attributes as measured by the log cross-product ratio is the same across sub-populations at each time point, then the $\{\Delta_{igk_1k_2}\}$ satisfy the hypothesis

$$H_{SA}: \ \Delta_{1gk_1k_2} = \Delta_{2gk_1k_2} = \ldots = \Delta_{sgk_1k_2} \quad \text{for } \begin{array}{l} g = 1,2,\ldots,t \\ k_1 = 1,2,\ldots,L_1-1 \\ k_2 = 1,2,\ldots,L_2-1 \end{array} , \quad (III.16)$$

where SA denotes sub-population association;

(11) If there is no sub-population x time interaction with respect to the log cross-product ratio measure of association, then the $\{\Delta_{igk_1k_2}\}$ may be written in terms of an additive model

$$H_{STA}: \ \Delta_{igk_1k_2} = \mu_{k_1k_2} + \xi_{i*k_1k_2} + \tau_{*gk_1k_2} \quad \text{for} \ \begin{array}{l} i = 1,2,\ldots,s \\ g = 1,2,\ldots,t \\ k_1 = 1,2,\ldots,L_1-1 \\ k_2 = 1,2,\ldots,L_2-1 \end{array}, \qquad \text{(III.17)}$$

where STA denotes sub-population x time association, and where $\mu_{k_1k_2}$ is an overall mean effect, $\xi_{i*k_1k_2}$ is a sub-population effect, and $\tau_{*gk_1k_2}$ is a time effect, and where it is understood that the $\{\mu_{k_1k_2}\}$, $\{\xi_{i*k_1k_2}\}$ and $\{\tau_{*gk_1k_2}\}$ satisfy the usual analysis of variance constraints. For an application of this type of additive model to measures of association, see Grizzle and Williams (1972).

Moreover, if the d attributes are all measured on the same L-point scale, hypotheses directed at the extent to which individual subjects are classified into the same category for each attribute can be investigated. For example, agreement on the classification of developmental stages by several different criteria is of considerable importance in establishing certain theories of behavioral growth (see Wohlwill, 1973). These problems are similar to those raised in the general area concerned with the measurement of agreement, and as such have received attention in a wide range of research areas as reviewed recently in Landis and Koch (1975a, 1975b). In this regard, numerous measures of observer agreement have been proposed for categorical data, e.g., Goodman and Kruskal (1954), Cohen (1960, 1968), Fleiss (1971), Light (1971), and Cicchetti (1972). Most of these quantities are of the form

$$\kappa = \frac{\pi_o - \pi_e}{1 - \pi_e} \qquad \text{(III.18)}$$

where π_o is an observational probability of agreement and π_e is a hypothetical expected probability of agreement under an appropriate set of baseline constraints such as total independence of attribute classifications.

Furthermore, kappa-type measures of agreement directly analogous to (III.18) can be developed to investigate the joint agreement of several attributes, as well as the pairwise agreements of two selected attributes. In addition, sets of weights which reflect the role of each response profile in a given agreement index can be selected to investigate "path" models of development among several behavioral tasks as discussed in Wohlwill (1973). Applications of such generalized kappa-type measures of agreement to clinical diagnosis data involving several observers is discussed in Landis and Koch (1977a, 1977b). In particular, the choice of weights which are in a hierarchical relationship with each other can be used to investigate hypothesized patterns of development such as synchronous progression, convergent "decalage," divergent "decalage," and reciprocal interaction (see Wohlwill, 1973, p. 215).

C. Estimation and Hypothesis Testing

Test statistics for the hypotheses considered in the previous sections as well as estimators for corresponding model parameters can be obtained by using the general approach for the analysis of multivariate categorical data discussed by GSK (1969). This procedure can be implemented by constructing the appropriate functions of the observed proportions which are directed at the relationships under investigation by a sequence of matrix operations. Then a weighted least squares computational algorithm is used to generate linearized minimum modified chi-square test statistics. The basic elements of the GSK procedure which pertain to this paper are summarized in the Appendix.

All the hypotheses in Section III.A involving constraints on the first-order marginal probabilities can be tested by expressing the estimates of the $\{\phi_{igmk}\}$ or the $\{\eta_{igm}\}$ as linear functions of the type given in the Appendix (A.14). Although these particular matrix expressions have already been discussed in considerable detail in Koch and Reinfurt (1971) and Koch *et al*. (1977) they will be presented within the context of the data analysis in Section 4. Otherwise, their specific construction for hypotheses like (III.2)-(III.6) is also documented in Landis (1975).

In contrast to the linear functions which pertain to the hypotheses in Section III.A, all the hypotheses involving measures of association and agreement require the expression of the corresponding ratio estimates as compounded logarithmic-exponential-linear functions of the observed proportions as formulated in the Appendix (A.20, A.21). As a result, the test statistics for the hypotheses in Section III.B can also be generated by the corresponding expression given in the Appendix (A.11).

IV. ANALYSIS OF LONGITUDINAL DATA EXAMPLES

This section is concerned with the analysis of the longitudinal data from examples II.A and II.B presented in Tables 1 and 2 with primary emphasis given to illustrating the methodology in Section III. In this regard, tests of significance are used in a descriptive context to identify important sources of variation as opposed to a rigorous inferential context; thus issues pertaining to multiple comparisons are ignored here. These, however, can be handled by the Scheffé type procedures given in GSK (1969).

A. Analysis of One-Population Example

The comparisons required to answer the questions associated with the example in Section II.A can be described more clearly within the context of two sub-tables of Table 1 corresponding to the cross-classification of the two attributes at each time point as shown in Table 3.

Insert Table 3 About Here

This study involves $s = 1$ sub-population, $t = 2$ time points (T1 and T2), $d = 2$ attributes (A1 and A2), $L_1 = 2$ response categories for A_1 and $L_2 = 2$ response categories for A_2. Thus, there are $r = (L_1 L_2)^t = 4^2 = 16$ possible multivariate response profiles.

The functions required to test the hypotheses involving the first-order marginal distributions can be generated in the formulation of (A.14) by using

$$A_1 = \begin{bmatrix} 0000 & 0000 & 1111 & 1111 \\ 0000 & 1111 & 0000 & 1111 \\ 0011 & 0011 & 0011 & 0011 \\ 0101 & 0101 & 0101 & 0101 \end{bmatrix} . \qquad \text{(IV.1)}$$

This yields the function vector

$$\underset{\sim}{F}' = (0.398, 0.441, 0.729, 0.788), \qquad \text{(IV.2)}$$

which contains the occurrence rates of A1 and A2 at each of the time points as shown in Figure 1.

Insert Figure 1 About Here

Table 3

CROSS CLASSIFICATION OF ATTRIBUTE DATA BY TIME POINTS

Time Point		T1				T2		
Attribute		A2				A2		
	Category	1	2	TOTAL	Category	1	2	TOTAL
A1	1	180	33	213	1	57	39	96
	2	18	123	141	2	18	240	258
	TOTAL	198	156	354	TOTAL	75	279	354

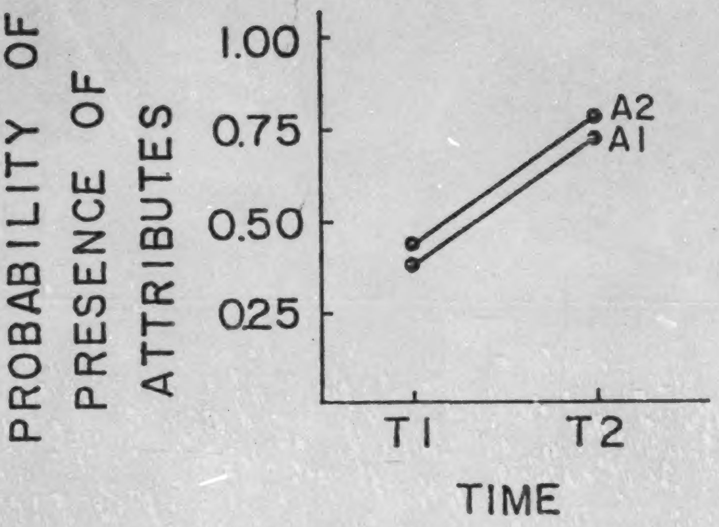


Figure 1-- Occurrence rates of two attributes (A1, A2) at each of two time points (T1, T2).

Consequently, the hypotheses associated with question (1)-(3) can be tested in the linear models phase of the analysis by setting $\underset{\sim}{X} = \underset{\sim}{I}_4$ and testing each of the following contrast matrices:

$$\underset{\sim}{C}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad ; \qquad \text{(IV.3)}$$

$$\underset{\sim}{C}_2 = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \quad ; \qquad \text{(IV.4)}$$

$$\underset{\sim}{C}_3 = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} \quad ; \qquad \text{(IV.5)}$$

$$\underset{\sim}{C}_4 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \quad ; \qquad \text{(IV.6)}$$

$$\underset{\sim}{C}_5 = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix} \quad ; \qquad \text{(IV.7)}$$

$$\underset{\sim}{C}_6 = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} \quad ; \qquad \text{(IV.8)}$$

$$C_7 = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \quad ; \qquad \text{(IV.9)}$$

The hypotheses from Section III which correspond to the $\underset{\sim}{C}$ matrices and the resulting test statistics are given in Table 4. These results suggest that significant differences ($\alpha = 0.05$) exist between the occurrence rates of the attributes at each of the time points, and that the occurrence rates of each attribute are significantly different ($\alpha = 0.01$) between the time points. Otherwise, the attribute x time interaction is not significant ($\alpha = 0.25$), which indicates the similarity of the change over time in the occurrence rates of the two attributes.

Insert Table 4 About Here

Table 4

TESTS OF HYPOTHESES INVOLVING MARGINAL DISTRIBUTIONS

Hypothesis	d.f.	$\underline{Q}_C$
H_{AMH}: ATTRIBUTES		
$\underset{\sim}{C}_1$ (T1 & T2)	2	13.32**
$\underset{\sim}{C}_2$ (T1)	1	4.47*
$\underset{\sim}{C}_3$ (T2)	1	7.91**
H_{TMH}: TIME		
$\underset{\sim}{C}_4$ (A1 & A2)	2	268.52**
$\underset{\sim}{C}_5$ (A1)	1	162.33**
$\underset{\sim}{C}_6$ (A2)	1	188.49**
H_{AT}: ATTRIBUTE X TIME		
$\underset{\sim}{C}_7$	1	0.32

* significant at $\alpha = 0.05$
** significant at $\alpha = 0.01$

The data from Table 1 can also be displayed in terms of the joint distribution of the two attributes cross-classified by the time points as shown in Table 5. In this context, the bivariate distributions of the two attributes for T1 and T2 are summarized in the row and column margins respectively. This joint distribution can be generated for each time point in the formulation of (A.14) by using

$$\underset{\sim}{A}_1 = \begin{bmatrix} 1111 & 0000 & 0000 & 0000 \\ 0000 & 1111 & 0000 & 0000 \\ 0000 & 0000 & 1111 & 0000 \\ 1000 & 1000 & 1000 & 1000 \\ 0100 & 0100 & 0100 & 0100 \\ 0010 & 0010 & 0010 & 0010 \end{bmatrix} . \qquad \text{(IV.10)}$$

Insert Table 5 About Here

Then by setting $\underset{\sim}{X} = \underset{\sim}{I}_6$, the hypothesis H_{TJD} in (III.13) associated with question (4) concerning differences between the two time points can be tested by using

$$\underset{\sim}{C} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} . \qquad \text{(IV.11)}$$

For these data, the test statistic for H_{TJD} is $Q_{\underset{\sim}{C}} = 268.53$ with d.f. = 3, which implies significant differences ($\alpha = 0.01$) between the joint distributions of A1 and A2 at the two time points. In particular, we observe in Table 5 that the major difference in the bivariate distributions is the shift from the large proportion (180/354) of the subjects who had neither attribute present at T1 to the large proportion (240/354) who had both attributes present at T2. However, we also note that this shift was not attributable only to individual subjects moving directly from (11) to (22). In fact, this distributional change is due to the high probability (48/51) of subjects who had only one of the attributes

Table 5

JOINT DISTRIBUTION OF ATTRIBUTES CROSS CLASSIFIED BY TIME POINTS

Time Point	Attribute categories (A1, A2)	T2 11	12	21	22	TOTAL
T1	11	57	36	18	69	180
	12	0	0	0	33	33
	21	0	3	0	15	18
	22	0	0	0	123	123
	TOTAL	57	39	18	240	354

present at T1 moving to the state of having both attributes present at T2, together with the tendency for those who had neither attribute present at T1 to progress to having either one or both of the attributes present at T2.

Although observed frequencies of zero as displayed in Table 5 can cause computational problems in certain applications, their presence is not troublesome here. In principle, this table has 15 degrees of freedom, but in terms of the observed data there are effectively only 7 d.f. which can be manipulated (without computational singularities as discussed in the Appendix), unless certain zero cells are replaced by 0.5. However, the 6 functions associated with the bivariate distributions specified in (IV.10) do not require such artificial data adjustments; thus, they can be analyzed directly. For a more detailed discussion concerning the treatment of observed zeros, see Koch et al. (1977).

Furthermore, the measures of association and agreement between A1 and A2 in Section III can be generated as compounded functions of the underlying vector of proportions. In particular, for each of the two time points, the log cross-product ratio in (III.10) can be generated in the formulation of (A.20) by using

$$\underset{\sim}{A}_1 = \begin{bmatrix} 1111 & 0000 & 0000 & 0000 \\ 0000 & 1111 & 0000 & 0000 \\ 0000 & 0000 & 1111 & 0000 \\ 0000 & 0000 & 0000 & 1111 \\ 1000 & 1000 & 1000 & 1000 \\ 0100 & 0100 & 0100 & 0100 \\ 0010 & 0010 & 0010 & 0010 \\ 0001 & 0001 & 0001 & 0001 \end{bmatrix} ; \qquad \text{(IV.12)}$$

$$\underset{\sim}{A}_2 = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} ; \qquad \text{(IV.13)}$$

the measure of association Q in (III.11) can be estimated in the formulation of (A.21) by using $\underset{\sim}{A}_1$ in (IV.12);

$$\underset{\sim}{A}_2 = \begin{bmatrix} 1001 & 0000 \\ 0110 & 0000 \\ 0000 & 1001 \\ 0000 & 0110 \end{bmatrix} ; \qquad \text{(IV.14)}$$

$$\underset{\sim}{A}_3 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} ; \qquad \text{(IV.15)}$$

$$\underset{\sim}{A}_4 = \begin{bmatrix} 1 & -1 \end{bmatrix} \qquad \text{(IV.16)}$$

and finally, Cohen's kappa in (III.18) under the baseline constraints of independence can be computed in the formulation of (A.21) by letting

$$\underset{\sim}{A}_1 = \begin{bmatrix} 1111 & 1111 & 0000 & 0000 \\ 0000 & 0000 & 1111 & 1111 \\ 1111 & 0000 & 1111 & 0000 \\ 0000 & 1111 & 0000 & 1111 \\ 1111 & 0000 & 0000 & 1111 \\ 1100 & 1100 & 1100 & 1100 \\ 0011 & 0011 & 0011 & 0011 \\ 1010 & 1010 & 1010 & 1010 \\ 0101 & 0101 & 0101 & 0101 \\ 1001 & 1001 & 1001 & 1001 \end{bmatrix} ; \qquad \text{(IV.17)}$$

$$\underset{\sim}{A}_2 = \begin{bmatrix} 10100 & 00000 \\ 10010 & 00000 \\ 01100 & 00000 \\ 01010 & 00000 \\ 00001 & 00000 \\ 00000 & 10100 \\ 00000 & 10010 \\ 00000 & 01100 \\ 00000 & 01010 \\ 00000 & 00001 \end{bmatrix} ; \qquad \text{(IV.18)}$$

$$\underset{\sim}{A}_3 = \begin{bmatrix} -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} ; \qquad \text{(IV.19)}$$

$$\underset{\sim}{A}_4 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} . \qquad \text{(IV.20)}$$

The estimates of these measures of association and agreement between A1 and A2 for the data in Table 1, together with their estimated standard errors are displayed in Table 6. Furthermore, the difference between the two time points with respect to each of these measures of association can be tested individually by setting $\underset{\sim}{X} = \underset{\sim}{I}_2$ and $C = [1\ -1]$ for Δ, Q, and κ respectively. In this regard, the corresponding test

----

Insert Table 6 About Here

----

statistics for this hypothesis in (III.15) associated with question (5) are displayed in Table 7. Here we note that although the correlation

----

Insert Table 7 About Here

----

structure between A1 and A2 (as measured either by Δ or Q) did not change between T1 and T2, the agreement between A1 and A2 is significantly different ($\alpha = 0.05$) between the two time points. This decrease in the agreement statistic from 0.70 to 0.56 is due largely to the increase in the expected value for the presence of both attributes (22), without a corresponding increase in the observed proportion of overall agreement.

B. Analysis of Two-Population Example

The example in Section II.B involves s = 2 sub-populations (boys, girls), t = 3 time periods (year 1, year 2, year 4), d = 1 behavioral task, and L = 2 response categories (success S and failure F). Thus, there are $r = L^t = 2^3 = 8$ possible multivariate response profiles.

Table 6

MEASURES OF ASSOCIATION AND AGREEMENT BETWEEN A1 AND A2

Time Period	T1		T2	
	Estimate	Estimated Standard Error	Estimate	Estimated Standard Error
Δ	3.62	0.316	2.97	0.321
Q	0.95	0.016	0.90	0.030
κ	0.70	0.038	0.56	0.051

Table 7

TEST STATISTICS FOR TIME DIFFERENCES IN MEASURES OF ASSOCIATION AND AGREEMENT BETWEEN A1 AND A2

Hypothesis	d.f.	Q_C
$\Delta_1 = \Delta_2$	1	2.02
$Q_1 = Q_2$	1	1.76
$\kappa_1 = \kappa_2$	1	5.03*

* significant at $\alpha = 0.05$

In this regard, differences in the growth profiles for the boys and girls can be investigated by using

$$\underset{\sim}{A}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \otimes \underset{\sim}{I}_2 \qquad \text{(IV.21)}$$

to generate estimates for the first-order marginal probabilities of success (S) shown in Figure 2 and failure (F) for each time x sex group combination in the formulation of (A.14), where $\otimes$ denotes Kronecker product of matrices and $\underset{\sim}{I}_u$ is the u x u identity matrix.

Insert Figure 2 About Here

Although a straightforward profile analysis could be performed directly on these estimated marginal probabilities, we will illustrate an alternative approach involving an underlying logistic model which is often of interest in growth studies (e.g., see Kowalski and Guire, 1974, and Guire and Kowalski, this volume). These involve log ratios or logit functions which can be generated in the formulation of (A.20) by selecting $\underset{\sim}{A}_2 = [1\ -1] \otimes \underset{\sim}{I}_6$, together with $\underset{\sim}{A}_1$ in (IV.21). These estimated probabilities of success and their corresponding logits, together with their respective estimated standard errors are shown in Table 8.

Insert Table 8 About Here

For this analysis, let λ_{ig} denote the asymptotic expected value of the logit corresponding to the i-th sex and g-th year. If time

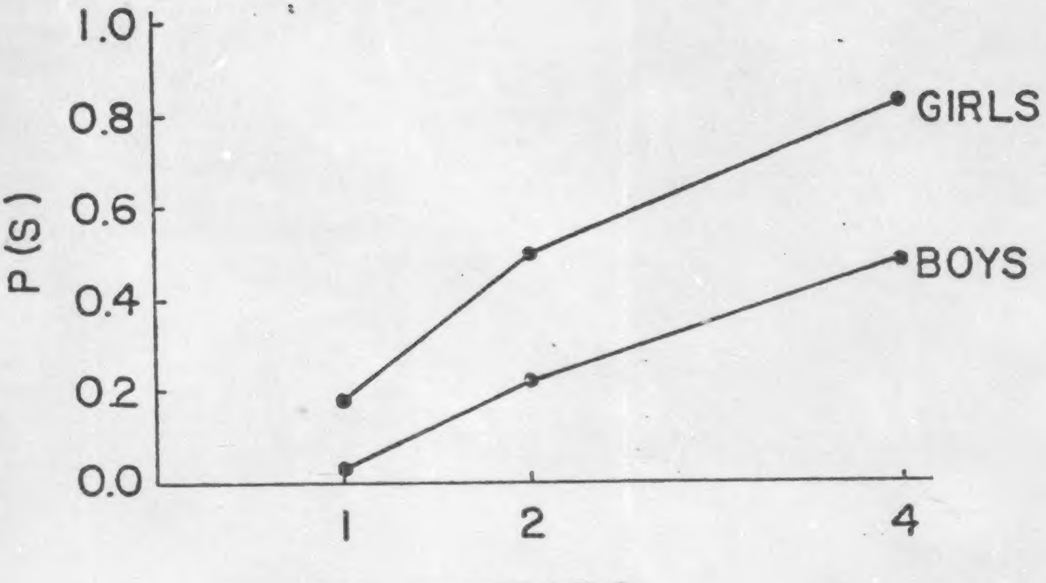


Figure 2-- Probability of success [P(S)] on a behavioral task for boys and girls at three different years of age.

Table 8

OBSERVED AND PREDICTED ESTIMATES FOR FIRST ORDER MARGINAL PROBABILITIES OF SUCCESS AND CORRESPONDING LOGITS

Sex Group	Year	Observed est. prob. success	Est. s.e.	Observed est. logit	Est. s.e.	Predicted est. logit	Est. s.e.	Predicted est. prob. success	Est. s.e.
Boys	1	0.03	0.01	-3.45	0.45	-2.99	0.20	0.05	0.01
Boys	2	0.22	0.03	-1.29	0.19	-1.50	0.14	0.18	0.02
Boys	4	0.48	0.04	-0.10	0.16	0.00	0.14	0.50	0.03
Girls	1	0.18	0.03	-1.50	0.18	-1.48	0.14	0.19	0.02
Girls	2	0.50	0.03	-0.01	0.14	0.01	0.09	0.50	0.02
Girls	4	0.83	0.03	1.57	0.18	1.51	0.14	0.82	0.02

is assumed to represent a metric which is analogous to drug dosage in quantal bioassay research, then the linear logistic model with respect to log time represents a reasonable model by analogy to well known results discussed by Berkson (1944, 1953, 1955) or Finney (1964). More specifically, we first consider the model

$$\lambda_{ig} = \mu_i + \gamma_i x_{ig} \quad \text{for } \begin{array}{l} i = 1,2 \\ g = 1,2,3 \end{array} , \qquad \text{(IV.22)}$$

where μ_i represents an intercept parameter in reference to year 1 which is associated with the i-th sex, γ_i represents a corresponding continuous slope effect over time, and x_{ig} is the log to the base 2 of year g for the i-th sex. In matrix notation, this model can be fitted via the regression model

$$\underset{\sim}{E}_A\{\underset{\sim}{F}\} = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \gamma_1 \\ \mu_2 \\ \gamma_2 \end{bmatrix} , \qquad \text{(IV.23)}$$

for which the goodness of fit statistic is $Q = 2.29$ with d.f. = 2. The hypotheses and test statistics in Table 9 suggest differences exist among the respective sex groups with respect to the intercept, but not the slope. On the basis of these results, the original

---

Insert Table 9 About Here

---

model can be simplified to

$$\underset{\sim}{E}_A\{\underset{\sim}{F}\} = \underset{\sim}{X}_2\underset{\sim}{\beta}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \gamma \end{bmatrix} \qquad \text{(IV.24)}$$

Table 9

STATISTICAL TESTS FOR $\underset{\sim}{X}_1$ MODEL

Hypothesis	d.f.	$\underset{\sim}{Q}_C$
$\mu_1 = \mu_2$	1	18.95**
$\gamma_1 = \gamma_2$	1	0.22

**significant at $\alpha = 0.01$

where μ_i is the intercept parameter for the i-th sex group and γ is an overall slope parameter. For this model, the goodness of fit statistic is $Q = 2.51$ with d.f. = 3, which suggests that this reduced model provides a satisfactory characterization of the variation among the logits. The corresponding estimated parameter vector $\underset{\sim}{b}_2$ and its estimated covariance matrix $V_{\underset{\sim}{b}_2}$ are given in (IV.25).

$$\underset{\sim}{b}_2 = \begin{bmatrix} -2.99 \\ -1.48 \\ 1.50 \end{bmatrix} ; \quad V_{\underset{\sim}{b}_2} = \begin{bmatrix} 3.99 & & \\ 1.60 & 1.94 & \\ -1.60 & -1.07 & 1.07 \end{bmatrix} \times 10^{-2} \qquad \text{(IV.25)}$$

From these results, the predicted logits shown in Table 8 can be determined via (A.12). These can then be used to obtain the predicted values for the first-order marginal probabilities of success (S) responses by reverse transformation which are illustrated in considerably more detail in Koch *et al.* (1977) and Landis *et al.* (1976). These quantities are also shown in Table 8 and are plotted in Figure 3 within the context of fitted logistic curves. Estimated standard errors for these predicted values obtained through suitable manipulations of (A.13) are substantially smaller than those for the corresponding observed estimates, and thus reflect the extent to which the fitted model $\underset{\sim}{X}_2$ enhances statistical efficiency.

Insert Figure 3 About Here

Finally, it can be shown that for this linear logistic model the parametric functions $\exp(-\mu_i/\gamma)$ represent the median ages for successful performance of the behavioral task (the ED-50 analogue from bioassay studies) in the i-th sex group. Estimates for these quantities are obtained as corresponding functions of $\underset{\sim}{b}_2$ and are shown in Table 10.

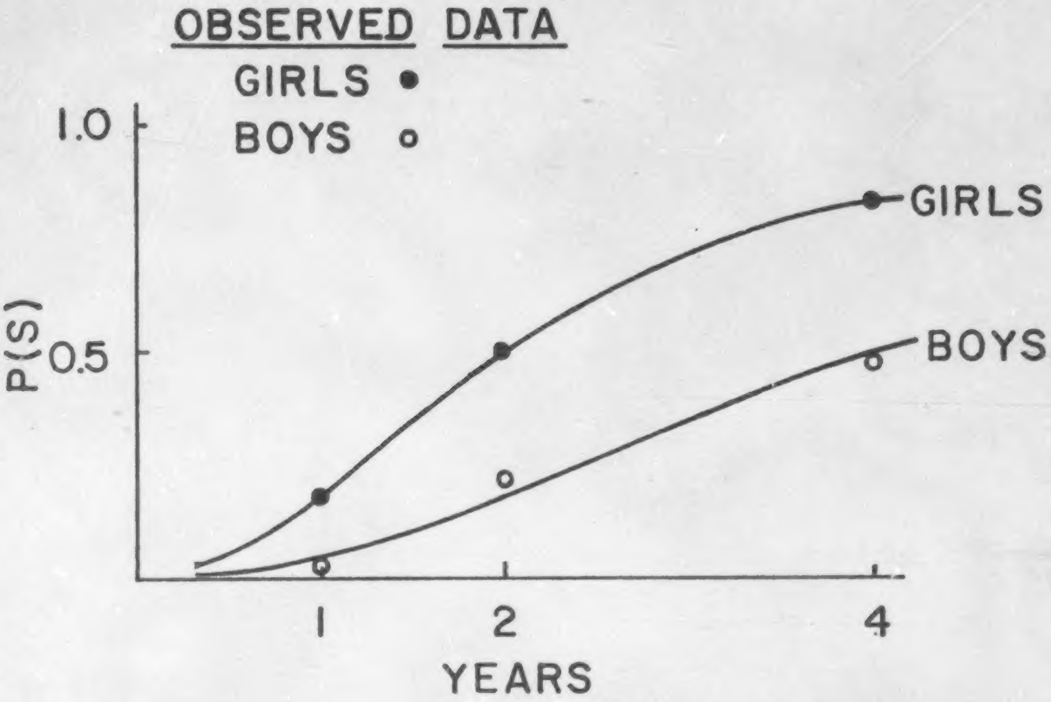


Figure 3 -- Fitted logistic curves to probability of success on a behavioral task [P(S)] for boys and girls at three different years of age.

Insert Table 10 About Here

V. DISCUSSION

Although the methodology for the analysis of longitudinal data developed in this paper is quite general, these procedures have been illustrated with relatively simple examples. However, for situations in which either the number of time points t, the number of attributes d, or the number of categories $L_1, L_2, \ldots, L_d$ are moderately large, the number of possible multivariate response profiles $r = (L_1 L_2 \ldots L_d)t$ becomes extremely large. Consequently, the matrices required to implement the GSK procedures directly may be outside the scope of computational feasibility. In addition, for each of the s sub-populations many of the r possible response profiles will not necessarily be observed in the respective samples so that corresponding cell frequencies are zero. Thus, in such cases, specialized computing procedures are required to obtain the estimates of the pertinent functions.

One alternative approach for handling such very large contingency tables in which most of the observed cell frequencies are zero is discussed in Koch et al. (1977) and is illustrated in Landis and Koch (1977b). Specifically, this approach permits the same estimators which would need to be obtained from the conceptual multidimensional contingency table to be generated by first forming appropriate indicator variables of the raw data from each subject, and then computing the across-subject arithmetic means. Subsequent to these preliminary steps, the usual matrix operations discussed in the Appendix can then be applied to these indicator variable means to generate the required

Table 10

ESTIMATED MEDIAN AGE FOR SUCCESSFUL PERFORMANCE OF BEHAVIORAL TASK

Sex Group	Estimated Median Age	Estimated Standard Error
Boys	4.00	0.25
Girls	1.99	0.09

functions. These alternative computations involving raw data, as well as these involving standard contingency table data, can all be performed via the computer program GENCAT discussed in Landis _et al_. (1976).

Otherwise, several additional potential difficulties associated with the introductory remarks in Section III may arise in the analysis of longitudinal data in studies of human growth and development. First of all, if the response profiles are assumed to be structured (e.g., as specified by certain irreversible growth or learning patterns as discussed in Wohlwill, 1973), some of the $\pi_{i\underline{j}}$ will be zero. In such situations, analogous hypotheses to those discussed in Section III that reflect these restrictions imposed by such structures can be taken into account by suitably modifying the definition of the appropriate hypotheses. In particular, the hypotheses pertaining to the first-order marginal probabilities are still appropriate here because their formulation is consistent with the available degrees of freedom. However, the higher order margins involving joint distributions or measures of association may not be feasible depending specifically on the nature of the restrictions involved. Secondly, the methods discussed in Koch, Johnson, and Tolley (1972) represent a life table approach to dealing with the issues involved with subjects who are lost-to-follow-up. Finally, for a discussion of one approach to the analysis of incomplete data resulting from the failure to measure each subject on each attribute at each time point, see Koch, Imrey, and Reinfurt (1972).

SUMMARY

In this paper we have proposed an extremely general approach to the analysis of multivariate categorical data associated with

longitudinal investigations of human growth and development. For purposes of illustration, two hypothetical data sets were presented to indicate the range of statistical issues of interest in such studies, and the types of functions from corresponding multidimensional contingency tables which can be used to suggest answers to these questions. Within this context, a general methodology for the analysis of categorical data resulting from longitudinal studies was then developed in terms of specific hypotheses. In particular, hypotheses directed at questions pertaining to average differences among sub-populations, attributes, and time points were expressed in terms of constraints on the probabilities associated with first-order marginal distributions of the response profiles. Furthermore, hypotheses directed at relationships among the attributes at a given time point and the extent to which those relationships change across time were formulated in terms of comparisons among joint distributions and in terms of measures of association across time periods.

A general unifying approach to the analysis of multivariate categorical data was recommended to create test statistics for these hypotheses as well as estimators for corresponding model parameters. This procedure can be implemented by constructing the appropriate functions of the observed proportions which are directed at the relationships under investigation by a sequence of matrix operations. Then a weighted least squares computational algorithm is used to generate linearized minimum modified chi-square statistics as discussed in more detail in the Appendix.

An extensive analysis of each of the two data sets was presented within the context of the hypotheses of interest. In this regard, the matrices required to generate the appropriate functions and the matrices used to compute the test statistics were all specified in detail. Furthermore, the fitting of final smoothed models was illustrated in the second example in terms of an underlying linear logistic model often considered in growth studies. In particular, this modeling permitted the estimation of predicted values for the first-order marginal probabilities, together with their estimated standard errors, even though the analysis was performed on the logit transform scale.

Finally, this paper concludes with a discussion of certain computational difficulties associated with very large contingency tables, together with other potential difficulties associated with structured response profiles, missing data, and incomplete data.

Appendix

Let $j = 1,2,\ldots,r$ index a set of categories which correspond to the r possible response profiles associated with the simultaneous classifications of the subjects on the d attributes. Similarly, let $i = 1,2,\ldots,s$ index a set of categories which correspond to distinct sub-populations as defined in terms of pertinent independent variables. If samples of size n_i where $i = 1,2,\ldots,s$ are independently selected from the respective sub-populations, then the resulting data can be summarized in an $(s \times r)$ contingency table as shown in Table 11, where n_{ij} denotes the frequency of response category j in the sample from the i-th sub-population.

Insert Table 11 about here

The vector $\underset{\sim}{n}_i$ where $\underset{\sim}{n}_i' = (n_{i1}, n_{i2}, \ldots, n_{ir})$ will be assumed to follow the multinomial distribution with parameters n_i and $\underset{\sim}{\pi}_i' = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ir})$, where π_{ij} represents the probability that a randomly selected element from the i-th population is classified in the j-th response category. Thus, the relevant product multinomial model is

$$\phi = \prod_{i=1}^{s} \left\{ n_i! \prod_{j=1}^{r} \left[\pi_{ij}^{n_{ij}} \big/ n_{ij}! \right] \right\} \qquad \text{(A.1)}$$

with the constraints

$$\sum_{j=1}^{r} \pi_{ij} = 1 \quad \text{for } i = 1,2,\ldots,s. \qquad \text{(A.2)}$$

Table 11

OBSERVED CONTINGENCY TABLE

Sub-population	Response profile categories				
	1	2	...	r	Total
1	n_{11}	n_{12}	...	n_{1r}	n_1
2	n_{21}	n_{22}	...	n_{2r}	n_2
...	...	...	...	...	...
s	n_{s1}	n_{s2}	...	n_{sr}	n_s

Let $\underset{\sim}{p}_i = (\underset{\sim}{n}_i/n_i)$ be the (r x 1) vector of observed proportions associated with the sample from the i-th sub-population and let $\underset{\sim}{p}$ be the (sr x 1) compound vector defined by $\underset{\sim}{p}' = (\underset{\sim}{p}_1', \underset{\sim}{p}_2', \ldots, \underset{\sim}{p}_s')$. Thus, the vector $\underset{\sim}{p}$ is the unrestricted maximum likelihood estimator of $\underset{\sim}{\pi}$ where $\underset{\sim}{\pi}' = (\pi_1', \pi_2', \ldots, \pi_s')$. A consistent estimator for the covariance matrix of $\underset{\sim}{p}$ is given by the (sr x sr) block diagonal matrix $\underset{\sim}{V}(\underset{\sim}{p})$ with the matrices

$$\underset{(r\times r)}{\underset{\sim}{V}_i}(\underset{\sim}{p}_i) = \frac{1}{n_i}\left[D_{\underset{\sim}{p}_i} - \underset{\sim}{p}_i\underset{\sim}{p}_i'\right] \tag{A.3}$$

for i = 1,2,...,s on the main diagonal, where $D_{\underset{\sim}{p}_i}$ is an (r x r) diagonal matrix with elements of the vector $\underset{\sim}{p}_i$ on the main diagonal.

Let $F_1(\underset{\sim}{p}), F_2(\underset{\sim}{p}), \ldots, F_u(\underset{\sim}{p})$ be a set of u functions of $\underset{\sim}{p}$ which pertain to some aspect of the relationship between the distribution of the response profiles and the nature of the sub-populations. Each of these functions is assumed to have continuous partial derivatives through order two with respect to the elements of $\underset{\sim}{p}$ within an open region containing $\underset{\sim}{\pi} = \underset{\sim}{E}\{\underset{\sim}{p}\}$. If $\underset{\sim}{F} \equiv \underset{\sim}{F}(\underset{\sim}{p})$ is defined by

$$\underset{\sim}{F}' = [\underset{\sim}{F}(\underset{\sim}{p})]' = [F_1(\underset{\sim}{p}), F_2(\underset{\sim}{p}), \ldots, F_u(\underset{\sim}{p})] \ , \tag{A.4}$$

then a consistent estimator for the covariance matrix of $\underset{\sim}{F}$ is the (u x u) matrix

$$\underset{\sim}{V}_{\underset{\sim}{F}} = \underset{\sim}{H}[\underset{\sim}{V}(\underset{\sim}{p})]\underset{\sim}{H}' \tag{A.5}$$

where $\underset{\sim}{H} = [d\underset{\sim}{F}(\underset{\sim}{x})/d\underset{\sim}{x} \mid \underset{\sim}{x} = \underset{\sim}{p}]$ is the (u x sr) matrix of first partial derivatives of the functions $\underset{\sim}{F}$ evaluated at $\underset{\sim}{p}$. In all applications, the functions comprising F are chosen so that V_F is asymptotically non-singular.

The function vector F is a consistent estimator of $F(\pi)$. Hence, the variation among the elements of $F(\pi)$ can be investigated by fitting linear regression models by the method of weighted least squares. This phase of the analysis can be characterized by writing

$$E_A \{F\} \equiv E_A \{F(p)\} = F(\pi) = X\beta \qquad (A.6)$$

where X is a pre-specified (u x t) design (or independent variable) matrix of known coefficients with full rank $t \leq u$, β is an unknown (t x 1) vector of parameters, and "E_A" means "asymptotic expectation."

An appropriate test statistic for the goodness of fit of the model (A.6) is

$$Q = Q(X,F) = (RF)' \; [R \, V_F \, R']^{-1} \; RF, \qquad (A.7)$$

where R is any full rank [(u-t) x u] matrix orthogonal to X. Here Q is approximately distributed according to the χ^2 distribution with d.f. = (u-t) if the sample sizes $\{n_i\}$ are sufficiently large that the elements of the vector F have an approximate multivariate normal distribution as a consequence of Central Limit Theory (CLT). Test statistics such as Q are known as generalized Wald (1943) statistics and various aspects of their application to a broad range of problems involving the analysis of multivariate categorical data are discussed in Bhapkar and Koch (1968a, 1968b) and Grizzle et al. (1969).

However, these test statistics like (A.7) are obtained in actual practice by using weighted least squares as a computational algorithm which is justified on the basis of the fact that Q of (A.7) is identically equal to

$$Q = (\underset{\sim}{F}-\underset{\sim}{X}\underset{\sim}{b})' \, \underset{\sim}{V}_F^{-1} \, (\underset{\sim}{F}-\underset{\sim}{X}\underset{\sim}{b}), \qquad (A.8)$$

where $\underset{\sim}{b} = (\underset{\sim}{X}' \, \underset{\sim}{V}_F^{-1} \, \underset{\sim}{X})^{-1} \, \underset{\sim}{X}' \, \underset{\sim}{V}_F^{-1} \, \underset{\sim}{F}$ is a BAN estimator for $\underset{\sim}{\beta}$ based on the linearized modified χ_1^2 -statistic of Neyman (1949). In view of this identity demonstrated in Bhapkar (1966), both Q and $\underset{\sim}{b}$ are regarded as having reasonable statistical properties in samples which are sufficiently large for applying CLT to the functions $\underset{\sim}{F}$. As a result, a consistent estimator for the covariance matrix of $\underset{\sim}{b}$ is given by

$$\underset{\sim}{V}_b = (\underset{\sim}{X}'\underset{\sim}{V}_F^{-1}X)^{-1} \quad . \qquad (A.9)$$

If the model (A.6) does adequately characterize the vector $\underset{\sim}{F}(\underset{\sim}{\pi})$, tests of linear hypotheses pertaining to the parameters $\underset{\sim}{\beta}$ can be undertaken by standard multiple regression procedures. In particular, for a general hypothesis of the form,

$$H_0\text{:} \quad \underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{0} \ , \qquad (A.10)$$

where $\underset{\sim}{C}$ is a known (c x t) matrix of full rank $c \leq t$ and $\underset{\sim}{0}$ is a (c x 1) vector of 0's, a suitable test statistic is

$$\underset{\sim}{Q}_C = (\underset{\sim}{C}\underset{\sim}{b})' \, [\underset{\sim}{C}(\underset{\sim}{X}'\underset{\sim}{V}_F^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{C}']^{-1}\underset{\sim}{C}\underset{\sim}{b} \qquad (A.11)$$

which has approximately a χ^2-distribution with d.f. = c in large samples under H_0 in (A.10).

In this framework, the test statistic $\underset{\sim}{Q}_C$ reflects the amount by which the goodness of fit statistic (A.8) would increase if the model

(A.6) were simplified (or reduced) by substitutions based on the additional constraints implied by (A.10). Thus, these methods permit the total variation within $\mathbf{F}(\boldsymbol{\pi})$ to be partitioned into specific sources and hence represent a statistically valid analysis of variance for the corresponding estimator functions $\mathbf{F}$.

Predicted values for $\mathbf{F}(\boldsymbol{\pi})$ based on the model (A.6) can be calculated from

$$\hat{\mathbf{F}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-1}\mathbf{F} \quad . \tag{A.12}$$

Thus, consistent estimators for the variances of the elements of $\hat{\mathbf{F}}$ can be obtained from the diagonal elements of

$$\mathbf{V}_{\hat{\mathbf{F}}} = \mathbf{X}(\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-1}\mathbf{X})^{-1}\mathbf{X}' \quad . \tag{A.13}$$

The predicted values $\hat{\mathbf{F}}$ not only have the advantage of characterizing essentially all the important features of the variation in $\mathbf{F}(\boldsymbol{\pi})$, but also represent better estimators than the original function statistics $\mathbf{F}$ since they are based on the data from the entire sample as opposed to its component parts. Moreover, they are descriptively advantageous in the sense that they make trends more apparent and permit a clearer interpretation of the relationship between $\mathbf{F}(\boldsymbol{\pi})$ and the variables comprising the columns of $\mathbf{X}$.

Although the formulation of $\mathbf{F}(\mathbf{p})$ can be quite general, Grizzle _et al_. (1969) and Forthofer and Koch (1973) demonstrated that a wide range of problems in categorical data analysis could be considered within the framework of a few specified classes of compounded logarithmic, exponential, and linear functions of the observed proportions. However, these functions are all special cases of a broad class of functions which can be expressed in terms of repeated applications of any sequence of the following matrix operations:

(i) Linear transformations of the type

$$\underset{\sim}{F}_1(\underset{\sim}{p}) = \underset{\sim}{A}_1\underset{\sim}{p} = \underset{\sim}{a}_1 , \tag{A.14}$$

where $\underset{\sim}{A}_1$ is a matrix of known constants;

(ii) Logarithmic transformations of the type

$$\underset{\sim}{F}_2(\underset{\sim}{p}) = \underset{\sim}{\log}_{\underset{\sim}{e}}(\underset{\sim}{p}) = \underset{\sim}{a}_2 , \tag{A.15}$$

where $\underset{\sim}{\log}_{\underset{\sim}{e}}$ transforms a vector to the corresponding vector of natural logarithms;

(iii) Exponential transformations of the type

$$\underset{\sim}{F}_3(\underset{\sim}{p}) = \underset{\sim}{\exp}(\underset{\sim}{p}) = \underset{\sim}{a}_3 \tag{A.16}$$

where $\underset{\sim}{\exp}$ transforms a vector to the corresponding vector of exponential functions, i.e., of antilogarithms.

Then the linearized Taylor-series-based estimate of the covariance matrix of $\underset{\sim}{F}_k$ for k = 1,2,3, is given by (A.5), where the corresponding $\underset{\sim}{H}_k$ matrix operator is

$$\underset{\sim}{H}_1 = \underset{\sim}{A}_1 ; \tag{A.17}$$

$$\underset{\sim}{H}_2 = \underset{\sim}{D}_{\underset{\sim}{p}}^{-1} ; \tag{A.18}$$

$$\underset{\sim}{H}_3 = \underset{\sim}{D}_{\underset{\sim}{a}_3} ; \tag{A.19}$$

where $\underset{\sim}{D}_{\underset{\sim}{y}}$ is a diagonal matrix with elements of the vector $\underset{\sim}{y}$ on the main diagonal.

The hypotheses involving marginal distributions can all be tested in terms of linear functions of the form given in (A.14). Furthermore, log-linear functions of the form

$$\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{A}_2 \{ \underset{\sim}{\log}_{\underset{\sim}{e}}[\underset{\sim}{A}_1\underset{\sim}{p}]\} \qquad (A.20)$$

can be used to generate logits and log cross-product ratios; whereas compounded functions of the form

$$\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{\exp}[\underset{\sim}{A}_4(\underset{\sim}{\log}_{\underset{\sim}{e}}\{\underset{\sim}{A}_3[\underset{\sim}{\exp}(\underset{\sim}{A}_2\{\underset{\sim}{\log}_{\underset{\sim}{e}}[\underset{\sim}{A}_1\underset{\sim}{p}]\})]\})]\} \qquad (A.21)$$

can be used to generate complex ratio estimates such as Yule's Q statistic or generalized kappa-type statistics. As a result, the linearized Taylor-series-based estimates of the covariance matrices associated with $\underset{\sim}{F}(\underset{\sim}{p})$ in (A.20) and (A.21) can be obtained by repeated application of the chain rule for matrix differentiation. In particular, let

$$\underset{\sim}{a}_1 = \underset{\sim}{A}_1 p \; ; \qquad (A.22)$$

$$\underset{\sim}{a}_2 = \underset{\sim}{\exp}\{\underset{\sim}{A}_2[\underset{\sim}{\log}_{\underset{\sim}{e}}(\underset{\sim}{a}_1)]\} \; ; \qquad (A.23)$$

$$\underset{\sim}{a}_3 = \underset{\sim}{A}_3\underset{\sim}{a}_2 \; ; \qquad (A.24)$$

$$\underset{\sim}{a}_4 = \underset{\sim}{\exp}\{\underset{\sim}{A}_4[\underset{\sim}{\log}_{\underset{\sim}{e}}(\underset{\sim}{a}_3)]\} \, . \qquad (A.25)$$

Then the results in (A.17)-(A.19) can be used to provide a consistent estimate of the covariance matrix via (A.5) for $\underset{\sim}{F}(\underset{\sim}{p})$ in (A.20) by using

$$\underset{\sim}{H} = \underset{\sim}{A}_2 \underset{\sim}{D}^{-1}_{\underset{\sim}{a}_1} \underset{\sim}{A}_1 \, , \qquad (A.26)$$

and for $\underset{\sim}{F}(\underset{\sim}{p})$ in (A.21) by using

$$\underset{\sim}{H} = \underset{\sim}{D}_{\underset{\sim}{a}_4} \underset{\sim}{A}_4 \underset{\sim}{D}^{-1}_{\underset{\sim}{a}_3} \underset{\sim}{A}_3 \underset{\sim}{D}_{\underset{\sim}{a}_2} \underset{\sim}{A}_2 \underset{\sim}{D}^{-1}_{\underset{\sim}{a}_1} \underset{\sim}{A}_1 \, . \qquad (A.27)$$

Finally, Koch, Imrey, Freeman, and Tolley (1976) discuss the application of this general approach to implicitly defined functions of $\underset{\sim}{p}$ in the context of estimated parameters from fitted log-linear models. Thus, all aspects of this methodology can be directed at implicit functions which are based on maximum likelihood estimation equations corresponding to preliminary or intermediate (as opposed to final) models with a priori assumed validity; in other words, models in which the likelihood (A.1) initially (i.e., prior to any data analysis) satisfies both (A.2) as well as certain other constraints analogous to (A.6).

For purposes of completeness, it should be noted that other statistical procedures for the analysis of categorical data from longitudinal and other types of repeated measurement experiments are available in the literature. In this regard, Bishop, Fienberg, and Holland (1975, Chapter 8) discuss the application of maximum likelihood methods to test hypotheses of total symmetry and marginal symmetry as well as certain other hypotheses of interest. They also provide a relatively complete literature review of other papers dealing with similar questions including the early work of Bowker (1948).

CAUSAL MODELS IN LONGITUDINAL RESEARCH

David Rogosa

Stanford University

Table of Contents

CAUSAL MODELS IN LONGITUDINAL RESEARCH

David Rogosa

Stanford University

Abstract

I. Introduction

II. Logic of Causal Models

III. Construction of Causal Models

IV. Longitudinal Panel Designs

V. Summary

Acknowledgment

I am grateful to Frank Capell, Linda Martinson, and Noreen Webb for their comments and insights.

ABSTRACT

The use of causal models in longitudinal research is discussed with emphasis on their logic and construction. Advantages of longitudinal designs over cross-sectional designs for making causal statements are presented first. Following an argument for theory-based research, the usefulness of causal models for incorporating substantive theory and knowledge into the data analysis model is stressed. The construction of causal models from substantive examples is explained and illustrated as a two-stage process involving (1) the structural model which specifies the relations of the important constructs, and (2) the measurement model which relates the unobserved constructs to their observable measures. Longitudinal panel designs are considered extensively, and causal models are constructed from two substantive investigations conforming to a panel design. The use of correlations in these studies to support causal statements is shown to be misleading.

CAUSAL MODELS IN LONGITUDINAL RESEARCH

INTRODUCTION

In this chapter a set of methods for making causal inferences from longitudinal data will be considered. It is useful to distinguish between explanation (which implies some form of causal statement) and description in investigations. A descriptive approach addresses such questions as "how" and "how much", while an explanatory approach treats the question "why" and thus incorporates causal inference. (See the discussion of description vs. explanation in Wold, 1956). For example, a measurement of a child's cognitive functioning is a description, while an investigation of the dependence of cognitive functioning on maternal nurturance and the child's motivation is an example of explanation. (Clearly, explanation and causal inference are appropriate when a researcher is interested in the mechanism which generated the observable relation between variables). Explanation of developmental processes is the primary focus of this chapter, but the discussion has clear implications for a variety of applications to other content domains.

A. Why use longitudinal data for causal inferences?

Cross-sectional analysis was originally conceived as a practical and appropriate means to study longitudinal change. However, sharp differences between the findings of cross-sectional and longitudinal studies have led many researchers to doubt the utility of the former:

> Because of the striking discrepancies in the results of cross-sectional and longitudinal investigations (Damon, 1965; Kuhlen, 1963), developmental psychologists have felt the need to formulate more sophisticated models.
>
> (Labouvie, Bartsch, Nesselroade, & Baltes, 1974, p. 288)

An examination of the assumptions underlying cross-sectional designs explains the discrepancies. Cross-sectional analysis assumes that interindividual differences in development are stable over time. The measurement of certain individuals at one time period, then, would presumably yield the same results as would the measurement of other individuals at the same developmental stages at a different time. Coleman (1968) expressed this as an assumption of equilibrium:

> The cross-section analysis assumes, either implicitly or explicitly, that the causal processes have resulted in an equilibrium state. That is, the implicit assumption in regression analysis is that this is a stable relationship, which would give the same values for the regression coefficients in a later cross-section unless an exogenous [outside] factor disturbed the situation. (p. 444)

Cross-sectional research further assumes that the *inter*individual differences found between age groups can be interpreted as changes that an individual would undergo across time (*intra*individual change). It has been shown, however, that interindividual differences in growth do not adequately measure intraindividual growth, but rather reflect the increasing variability in the rate of growth among individuals (Huston-Stein & Baltes, in press). Thus, Baltes and Willis (1976) stated:

> change on the individual level involves *intraindividual change*, and differences in change functions between individuals [involve] *interindividual differences in change* .... One of the important features of aging is that interindividual differences typically increase with time and age, resulting in progressively less age-related homogeneity.
>
> (p. 12, italics in original)

Because cross-sectional studies confound intraindividual growth with interindividual differences in intraindividual growth (Baltes & Nesselroade, 1973), they are a poor method for studying change. These problems with cross-sectional designs have been noted in other research contexts:

> Repeated cross-sections from a panel of organizations which are out of equilibrium will ordinarily produce estimates varying considerably from cross-section to cross-section.... It should be apparent that, if the processes of study are not symmetric in growth and decline, cross-sectional analysis mixing growers and decliners will obscure the processes of interest.
>
> (Freeman & Hannan, 1975, p. 2[illegible]6)

Furthermore, Coleman (1968) argued that repeated measures designs are more useful than cross-sectional designs in demonstrating causality:

> When variables are observed at two or more points in time, additional information exists beyond that obtained in cross-sectional data. This is information which, if used properly, can indicate what factors bring about change in a variable. These changes will, of course, create or maintain the relationships that may be found in cross-sectional data, and thus provide information about the dynamics of a system beyond that provided by cross-sectional data. (p.445)

The use of temporal variation to establish causality is a major aspect of the rationale for the longitudinal panel designs in section IV.

Another advantage of longitudinal research lies in untangling the complex effects of reciprocal causation or what some have called "causal loops" (Hannan & Young, 1974). Examples of reciprocal causation are the reciprocal influences (of mother-on-child and child-on-mother) in a mother-child dyads (Lewis & Rosenblum, 1974; Rogosa & Ambron, 1976).

An example from research on personality development illustrates this point. A number of cross-sectional studies have shown that "there is no question that there is a persistent relationship between the self and academic achievement [and] that there is a continuous interaction between the self and academic acheivement, and that each directly influences the other" (Purkey, 1970, p. 23). In a longitudinal study of these variables, O'Mally and Bachman (1976) explained the advantages of such a design:

> There is reason to believe that self-esteem is linked to educational and occupational attainment, and that this linkage probably involves a number of different and complexly interrelated patterns of causation. While it is a rather simple matter to point out the complexities, the problem of disentangling such reciprocal causation is extremely difficult and vexing. In most cases the empirical evidence is limited to a static relationship at a single point in time--e.g., survey respondents with higher levels of educational attainment also have higher mean scores on a measure of self-esteem. While such findings are important in demonstrating that a relationship does exist, they leave us largely in the dark about causal dynamics. (p.5)

B. Longitudinal Research and Classical Experimental Design and Analysis

While longitudinal designs are preferable to cross-sectional ones in terms of demonstrating causality and studying intraindividual change, they also create practical problems. Most longitudinal studies must be performed outside of controlled laboratory settings, where the characteristic features of classical experimentation--control of extraneous influences, manipulation of treatment variables, and equivalent experimental groups--are frequently impossible to implement. Furthermore, developmental studies often focus on relations of variables that cannot be controlled or manipulated (e.g., organismic variables).

One advantage of this break with classical experimental design lies in the increased ecological validity of naturalistic studies which decreases the threat to the external validity or generalizability of the study (Bracht & Glass, 1968). Bronfenbrenner (1974, 1976) argued persuasively for the necessity of ecological validity in developmental research and claimed that "much of developmental psychology is the science of the strange behavior of children in strange situations with strange adults for the briefest possible periods of time (1974, p. 3). Bronfenbrenner (1976) called for a new perspective in developmental research termed "the experimental ecology of education".

The increasing recognition of the need for naturalistic, longitudinal studies has led to a press for the development of appropriate design and analytic techniques. From the perspective of life-span psychology (Huston-Stein and Baltes, in press) stated that "most of the traditional, experimental design methods in the psychological sciences are ill-suited for the assessment of long-term chains and distal causes. Therefore, life-span researchers have pointed to the general usefulness of quasi-experimental designs" (p.11). Given the break in design strategies with traditional psychological experimentation, the analysis techniques successful in making causal inferences from experimental data are not likely to be useful for the analysis of causal patterns in naturalistic studies. Wiley and Hornik (1973) discussed this trend in design and analysis

> Sociological and social-psychological research workers have long been concerned with the attribution of causality to variables representing basic sociological or psychological concepts. Given the difficulty of variable manipulation in many real social settings, scientists have turned to statistical methodology rather than experimental techniques of investigation for help. (p.1).

Huston-Stein and Baltes (in press) speculated that "with more powerful methods of inferring causal relationships from naturalistic, correlational data, child developmentalists may be lured away from their heavy reliance on laboratory experiments" (p. 11). Herman Wold (1956) in a paper "Causal inference from observational data", indicated one path to "the more powerful methods":

> in the absence of experiments the statistical analysis has to be closely coordinated with subject-matter theory both in specifying the causal hypotheses and in testing them against other sources of knowledge. A crucial feature is randomization, which in experimental situations reduces the disturbing effect of uncontrolled variation. This device not being available in observational situations, it is a pertinent problem to what extent the disturbance factors should be taken into explicit account, and at this point strong reliance must fall upon subject-matter argument. (p. 31)

The attribution of causal effects from nonexperimental data can be accomplished through causal models which incorporate substantive knowledge into the analysis of data as a substitute for experimental controls. The process by which substantive knowledge is incorporated into the data analysis in a causal model through the construction of causal models from longitudinal studies is illustrated in sections III and IV.

II. LOGIC OF CAUSAL MODELS

A. Theory-based Research

A strong appeal for theory-based research was made by Suppes (1973):

> I think the time has come to call for a much deeper theoretical orientation of research in education in order thereby to increase its relevance. In many areas, the greatest limitation on research is not the absence of hard-data studies, but the absence of serious and sophisticated theory (p. 23)

Suppes rejects the "firm prejudices and soggy arguments" common in educational debate in favor of "well-put-together theories that have definite and precise assumptions and deductive consequences that bear on behavior and the way students learn" (p. 24). Suppes (1974) argues that theory goes beyond empiricism to explain the complexity of phenomena. He attacks the "triviality of bare empiricism" which at the extreme is the simple recording of facts that leads nowhere.

A similar criticism was voiced by Einhorn (1972):

> As methods and techniques get more complicated, the role of theory in research is being dangerously ignored in favor of purely empirical work that proceeds without so much as a hypothesis. Like Pirandello's characters in search of an author, many of today's researchers seem to have an assortment of techniques in search of a substantive problem. (p. 367).

In order to guide research, theory should be explicit about the relations of its components--that is, be translatable into empirically justifiable statements--since vague, verbal theories (which Suppes terms fantasies) will rarely help to focus research. One of Suppes' examples of fantasies in educational research is Piaget's concept of developmental states since it "operates in large theoretical terms and with little regard for detailed experimental investigation" (Suppes, 1973, p. 13).

Phillips and Kelly (1975) examined the "much-touted hierarchical theories of development":

> Hierarchical theories may be potentially useful in understanding human development. At the present time, however, the scientific status of such theories is obscure. In the flurry of experimental research on child development, it is not always clear what type of evidence would count as confirmation or refutation of such a theory, or indeed whether empirical research is relevant at all. These and

> related issues have not been totally ignored in the literature, but discussion has been hampered by the failure to draw some important distinctions and to make explicit certain underlying assumptions. (p. 352)

The strongest condemnation of obscure, vague theory is provided in Popper's (1972) discussion of the psychological theories of Freud and Alder.

> This does not mean that Freud and Adler were not seeing certain things correctly: I personally do not doubt that much of what they say is of considerable importance, and may well play its part one day in a psychological science which is testable. But it does mean that those "clinical observations" which analysts naively believe confirm their theory cannot do this anymore than the daily confirmations which astrologers find in their practice. And as for Freud's epic of the Ego, the Super-ego, and the Id, no substantially stronger claim to scientific status can be made for it than for Homer's collected stories from Olympus. These theories describe some facts, but in the manner of myths. They contain most interesting psychological suggestions, but not in a testable form. (p. 23)

Explicit statments of theory are essential for worthwhile theory-based research. Many developmental theories are stated ambiguously; consequently, they are often misinterpreted and are difficult for the researcher to test. Theorists use notions like causes, forces, systems, properties, schemes, and stages. The research must infer the relations of these terms from the incomplete specification provided by the theory and must select or develop measures that act as indicators for the unmeasured variables in the theory. Unfortunately there are substantial variations in the ways researchers interpret theory. The more ambiguously the theory is stated, the more confused are the results of empirical investigations.

The gap between theory and research is not uncommon in social science. Blalock (1963) described the problem in sociology and one approach to a solution:

> There can be little disagreement that in sociological theory we have made use of numerous concepts that have been vaguely defined... These are the 'big words' sociologists are often accused of using. Sometimes they are thrown into theoretical discussions with almost reckless abandon. In other instances they may be rather vaguely linked with measured indicators that are referred to as 'correlates', 'manifestations', or 'symptoms' of the underlying variables. Sometimes the 'big words' are utilized to provide explanations for empirical relationships, but upon detailed logical investigations we find that these theoretical explanations make much less sense than supposed... One possibility is to retain whatever vaguely defined concepts we may think will ultimately prove useful, while at the same time attempting to spell out exactly how we might link these theoretical concepts with specific measured variables. (p. 62)

The major problem for empirical research is that when theories include ambiguous concepts whose postulated causal relations are not well-specified, the resulting research is usually a collection of non-comparable studies which relate only vaguely to the original theory. Two examples are research on Piaget's theory of moral development and research on attachment (Ambron and Rogosa, 1975).

B. What are Causal Models?

A causal model is a representation of the postulated causal links between the variables of interest and is an explicit and quantitative statement of theory. Through the use of causal models verbal theories are recast in terms of the causal processes assumed to operate among the variables under consideration. The use of causal models forces the theorist and the experimenter to make explicit all causal assumptions in an internally consistent system. Clearly, this is a large step in overcoming the problems with verbal theories. As Duncan (1966), asserted "The great merit of the path scheme, then, is that it makes the assumptions explicit and tends to force the discussion to be at least internally consistent, so that mutually incompatable assumptions are not introduced surreptitiously into different parts of an argument extending over scores of pages" (p. 9).

In the formulation of a causal model, the important variables in the developmental process are first identified, for example, self-esteem, educational attainment, occupational aspiration. Then the causal links between these variables over time are specified. On psychological grounds some variables can be said to influence others; in other instances the causal link may be assumed not to exist. The variables to be included in the model and their postulated causal links may be obtained from the theoretical formulation of the problem.

Causal models have been discussed under a variety of names in a number of different literatures. Structural equation models is the term used most often in econometrics; path analysis was formulated by Wright in genetics and brought over into sociology as path analysis or as causal modeling. Goldberger (1972) provided an interesting history of the parallel but independent historical developments in econometrics and biometrics. Structural equation models are more general in formulation and in estimation techniques than path analysis but the logic is nearly identical. The models in this chapter are termed causal models or structural equation models, but almost all comments apply to the special case of path analysis also.

Causal models are regression-based procedures. The regression equations which compose a causal model are called structural regression equations. Systems of structural regression equations, in which each equation represents a causal link between variables, are distinguished from predictive regression equations which represent empirical associations with no special regard for whether the predictor (i.e., independent) variables are theoretically interesting.

Predictive regression, on the one hand, is concerned with identifying the best linear predictor of the dependent variable from a combination of the independent variables; any observable predictors that add to the explained variance, the squared multiple correlation (R^2), are utilized. The usual multiple regression techniques are examples of predictive regression. In structural regression, on the other hand, the concern is with the interrelationships of the theoretically important variables, not simply the predictability of one from the others.

In structural regression, the mechanism that generated the observations can be characterized in terms of more fundamental parameters. Goldberger (1973) proved that, in general, the coefficients in predictive regression will be a mixture of the structural parameters; hence a change in one structural parameter may change all the predictive regression coefficients. Therefore, the more fundamental structural parameters have the invariance and stability desired of scientific formulations. As Abraham Wald (1940) pointed out, "The knowledge of the structural relationship is essential for constructing any theory in the empirical sciences... in deducing laws from observations we have the task of estimating structural relationships" (p.300). Tukey (1954) concluded, "Almost any causal theory comes sooner or later to deal with structural regression rather than predictive regression" (p. 41).

Once a causal model is constructed, the set of causal links in the model are written as a set of structural regression equations. Estimating the parameters in the structural regressions (see Chapter , this volume) yields estimates of the causal influences between the variables and thus is a calculation of how change in one variable in the system will affect the other variables in the system. The results of these analyses depend on the theory which determined the structural equation model. Intuitively, the estimation techniques for these structural equation models decompose the observed association of the variables into direct and indirect effects. We are, in some sense, taking the correlation apart and examining its causal components. However, this decomposition of the observed covariation depends upon the particular causal ordering assumed to be valid by the researcher. The incorporation of relevant substantive knowledge into the regression analysis may be termed analyzing data *conditional on a theory*. From the theoretical and psychological conceptualization of the investigation, the observed association between the observed variables is specified. Then the regression analysis proceeds to supply estimates of causal parameters from the nonexperimental, correlational data. Of course, the numbers obtained are reasonably correct only if the substantive specification is adequate.

These causal modeling techniques *cannot prove causality*; they can help one choose between relevant causal hypotheses by ruling out those not conforming to the data. This is the logic of falsification (Popper, 1972). When theories are expressed as causal models, they are subject to rejection if contradicted by data.

C. Spurious Correlation: A Causal Interpretation

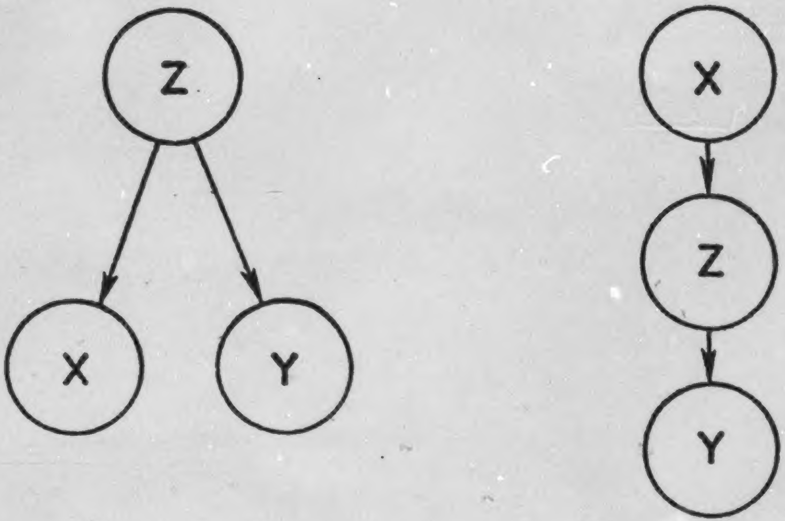
A specific problem which illustrates the importance of careful interpretation of the relationships between variables is that of spurious correlation, where the association between two variables is entirely due to the influence of a common factor. In investigating spurious correlation interest lies in whether a relation between two variables (x and y) disappears when a third variable z is introduced. The correlation of x and y is spurious if the association of x and y is totally due to the causal influence of z. This is illustrated in Figure 1a. To guard against this possibility we might compute the partial correlation $r_{xy.z}$ between x and y with z held constant. If $r_{xy.z}$ is effectively zero then this may be the spurious case in Figure 1a. But it is also possible that z is a mediating variable in the 'true' relationship of x and y as illustrated in Figure 1b.

Insert Figure 1 about here

Whether an association is spurious or "true" (from a causal standpoint) cannot be determined on the basis of correlations. Information about the causal ordering of the system of variables is required, which is best derived from substantive theory.

In Simon's (1954) classic example, a high negative correlation is found between x, the percentage of a group that is married and y, the average number of pounds of candy consumed per month per person. Can we conclude that marriages causes a reduction in candy consumption? Variable z is the average age of members in each of the several groups. However, when age is held constant, the correlation disappears. From common sense the relationship in Figure 1a is believed to hold; the correlation between candy consumption and marital status is jointly caused by a variation in age--the relationship is spurious. This is a 'common sense' conclusion, but it depends on the assumption that certain relations are not causal. In this example, the decision between models (a) and (b) was made by the a priori assumption that the age of a person does not depend upon marital status or candy consumption. Here the answer is obvious, but determining causal ordering and structure is often treacherous, and explicit statements of theory are necessary for the unambiguous interpretation of data.

Although problems with the causal ordering of variables affect all analysis schemes, a number of other characteristics of correlation coefficients make their use to support claims of causation in non-experimental studies



a. Correlation between X and Y is spurious since the association is entirely due to the causal influence of Z.

b. The true association between X and Y is mediated by Z.

Figure 1 -- Examples of true and spurious correlation. The partial correlation $r_{XY.Z}$ will equal zero in the population in both cases. Assumptions about the causal ordering of the variables are necessary to differentiate a from b.

inappropriate and often misleading. Correlation is a measure of the linearity of a relationship between two variables. Correlation is a descriptive statistic, and contains no information about the direction of the association. Thus correlations are inappropriate for causal statements. However, many longitudinal investigations report correlations as evidence of causal relationships. Longitudinal researchers may do well to heed the message of John Tukey (1954), who in an article on causal models, argued "correlation coefficients are justified in two and only two circumstances, when they are regression coefficients, or when the measurement of one or both variables on a determinate scale is hopeless". (p. 39)

In the substantive examples to be presented in Sections III and IV, some of the many pitfalls associated with the use of correlations for causal statements are illustrated. The use of regression coefficients solves some of the problem associated with the stability of reported effects over different samples. As Tukey (1954) noted, "We are very sure that the correlation cannot remain the same over a wide range of situations, but it is possible that the regression coefficients might" (p. 41). The regression coefficients that possess the desired stability are those of structural regression equations.

III. CONSTRUCTION OF CAUSAL MODELS

The construction of a causal model from a previously published developmental study is presented to clarify the preceding discussion of their properties. Radin (1971, 1974) investigated antecedents of cognitive development in lower-class children. Radin was primarily interested in the effects of child rearing practices, maternal behavior in particular. Based on theoretical and empirical studies Radin hypothesized that "maternal nurturance would foster intellectual functioning of the child" (1974, p. 1126). Radin also considered evidence that the motivation of the child to achieve is an intervening variable in the relationship between maternal behavior and cognitive functioning.

Radin's first hypothesis that maternal nurturance would foster the child's intellectual functioning can be represented as a relation between the constructs Maternal Behavior and Cognitive Functioning as depicted in Figure 2a. The arrow leading to Cognitive Functioning indicates the hypothesized causal influence of Maternal Behavior on Cognitive Functioning.

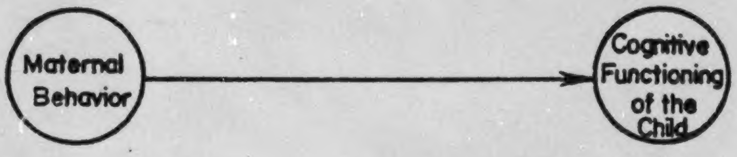
The consideration of motivation to achieve as a mediating variable can be incorporated into the relationship described in Figure 2a. Figure 2b includes the Motivation construct in the postulated causal ordering. The model in Figure 2b exhibits the two causal paths by which Maternal Behavior may influence Cognitive Functioning. The direct path connects Maternal Behavior and Cognitive Functioning, and the indirect path involves the Motivation construct as an intervening variable. Figure 2b is a representation of what is termed the structural model, which expresses the assumed relations between the unmeasured variables. Unmeasured or unobserved variables arise when the measurable variables differ from their theoretical counterparts. Unmeasured variables often are theoretical constructs which have implications for determining the relationship between observables. Since theory is most often expressed in terms of constructs and unmeasured variables, the structural model contains the statement of substantive knowledge to be incorporated into the data analysis. Causal models provide the machinery to relate theory and hypotheses expressed in unmeasured constructs to observed variables.

Frequently the variables actually measured are theoretically unimportant in their own right, but are taken as indicators of the underlying, theoretically-significant constructs. As Tukey (1954) explained "a prime characteristic of quantitative causal theories is that they include quantitative concepts which are not all subject to direct measurement" (p. 40). The goal is to make causal statements about the unmeasured variables from the observed relations of the indicators.

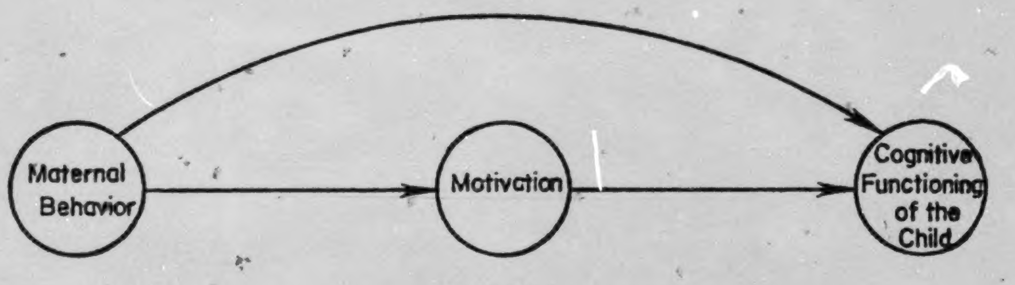
The measurement model provides the link between the constructs and the indicators. Radin measured two indicators for each of the constructs in Figure 2b. The two indicators of Maternal Behavior were measures of warmth and restrictiveness, obtained during a one hour observation of interaction between mother and child. The two indicators of motivation were part of the Pupil Behavioral Inventory (PBI) and a psychologist's rating of the child's motivation while taking the Stanford-Binet Intelligence Scale. For Cognitive Functioning the two indicators used were the Stanford-Binet and the Peabody Picture Vocabulary Test (PPVT).

Insert Figure 2 about here

Figure 3 depicts the full causal model, combining the structural and measurement models, of Radin's investigation. The symbols $\beta_1, \beta_2, \beta_3, \delta_1$ through δ_6) refer to the parameters of the causal paths which can be estimated from the data. In figure 3 the e_i represent the part of the observed variable that is not part of the construct it is presumed to measure. The e_i include measurement error and other information irrelevant to the construct. The e_i may be thought of as the unique part of the observed variable. Often the e_i are assumed to be uncorrelated with each other. In Figure 3 a correlation between e_4 and e_5 is allowed. This correlation is not assumed to be zero because X_4 and X_5, measures of Motivation and Cognitive Functioning respectively, are obtained on the same occasion, the administration of the Binet. It is reasonable to expect X_4 and X_5 to be related for reasons additional to the presumed relation between Motivation and Cognitive Functioning. The ability to estimate models with correlated errors is an important feature of the estimation techniques for these models (See Jöreskog & Sörbom, 1976a; and Chapter this volume).


Maternal Behavior
Cognitive Functioning of the Child

a. Representation of the hypothesized causal effect of Maternal Behavior on the Cognitive Functioning of the Child.


Maternal Behavior
Motivation
Cognitive Functioning of the Child

b. Representation of the three constructs (unmeasured variables) in the postulated causal ordering.

Figure 2 -- Representations of the causal relationship between Maternal Behavior and the Cognitive Functioning of the Child.

The causal model is equivalent to a series of regression equations with the same parameters as in Figure 3. The structural model is

$$
\begin{aligned}
X_B &= \beta_1 X_A + U \\
X_C &= \beta_2 X_B + \beta_3 X_A + V
\end{aligned}
$$

The measurement model is

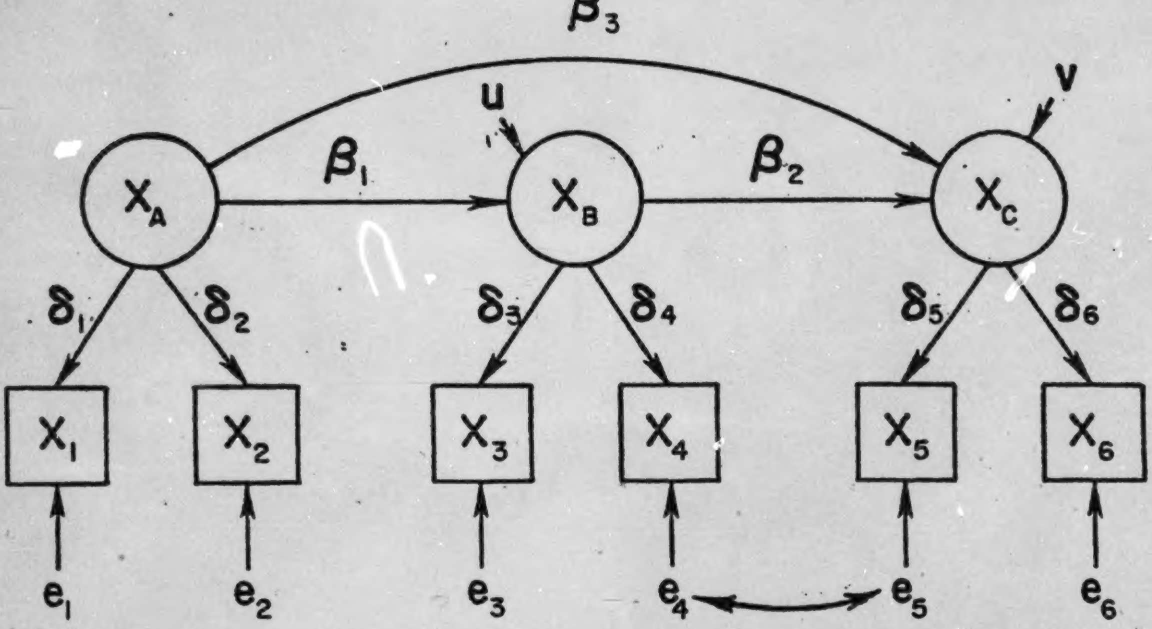
$$
\begin{aligned}
X_1 &= \delta_1 X_A + e_1 \\
X_2 &= \delta_2 X_A + e_2 \\
X_3 &= \delta_3 X_B + e_3 \\
X_4 &= \delta_4 X_B + e_4 \\
X_5 &= \delta_5 X_C + e_5 \\
X_6 &= \delta_6 X_C + e_6
\end{aligned}
$$

Insert Figure 3 about here

From these two sets of equations, relationships can be expressed strictly between the observables and the unknown parameters. These regression equations are called the reduced form. To estimate the model, the constraint $\delta_1 = \delta_3 = \delta_5 = 1$ is introduced, which constrains X_A to be in the same metric as X_1, etc. This constraint does not effect the generality of the analysis.

It should be noted that the usual multiple regression approach is inadequate when multiple measures (indicators) of the same construct are present. High collinearity leads to the problem of the 'bouncing beta weight' and results in theoretical nonsense (Gordon, 1968).

In this example the parameters β_1, β_2, and β_3 representing the causal influence between the constructs, are of central interest. The direct influence of Maternal Behavior on Cognitive Functioning is represented by β_3. The indirect effect of Maternal Behavior on Cognitive Functioning, with Motivation as the mediating variable, is the product of β_1 and β_2. The importance of Motivation


β_3
u
V
X_A
β_1
X_B
β_2
X_C
δ_1
δ_2
δ_3
δ_4
δ_5
δ_6
X_1
X_2
X_3
X_4
X_5
X_6
e_1
e_2
e_3
e_4
e_5
e_6

Constructs (unmeasured variables)

X_A – Maternal Behavior
X_B – Motivation of Child
X_C – Cognitive Functioning of the Child

Indicators (measured variables)

X_1 – Maternal Warmth
X_2 – Maternal Restrictiveness
X_3 – Motivation (PBI)
X_4 – Motivation during Binet
X_5 – Stanford Binet Score
X_6 – PPVT Score

Figure 3 -- Complete causal model of Radin's investigation with constructs and indicators labeled.

as a mediating variable can thus be gauged by the relative magnitudes of the estimates of β_3 and β_1, β_2.

Radin's original analysis (1971, Table 2) did not produce conclusive results on the importance of motivation. In the original analysis, using partial correlations between the indicators, all of the indicators could not be considered simultaneously. Causal model methods can estimate the relationships between the unmeasured variables using all the data at once. A reanalysis based on the present formulation (Rogosa, Webb, & Radin, 1976) indicated that motivation is a very important, intervening variable for the influence of Maternal Behavior on Cognitive Functioning.

The process of selecting the variables to be included in the model and postulating the causal relationships of the included variables and the residual terms constitutes the specification of the model. Mistakes in the specification such as omitting an important causal variable or incorrectly assuming that a causal path does not exist are termed specification errors. Duncan (1975a) suggests that specification error "is quite a useful euphemism for what in blunter language would be called 'using the wrong model'. There are many more wrong models than right ones, so that specification error is very common, though often not recognized and usually not easily recognizable" (p. 101).

In the Radin model, at least two measures of each construct were needed in order to estimate the parameters in the model. This is the technical problem of identification; a model is identified when all the causal parameters are uniquely estimable from the data. Underidentification results when there is not enough information to estimate certain causal parameters.

For many constructs used in developmental research a single measure will rarely be sufficient for valid measurement, and thus multiple indicators are important for both the technical concerns with identification and for the substantive problems of valid measurement. Baltes and Nesselroade (1973) argued:

> if one wishes to deal with more abstract concepts, such as patterned change--change in the interrelationships among a variety of measures--then a multivariate approach is necessary. It is from the interrelationships among measures (e.g., covariances), so often ignored in developmental research, that we may eventually extract the raw material that can be efficiently molded into general, but powerful constructs to aid the scientific study of development. A related point is that when research interest centers on change in more molar behavior patterns such as anxiety and aggression, no single variable can serve as a perfect indicator of the target construct. The use of multiple indicators (measures) enables us to form some combination of measures which "locates" the construct more precisely. (p. 222)

The choice of multiple indicators is not always desirable, since it indicates the presence of redundant information and not necessarily an increase in measurement validity. Detailed discussion of the choice and interpretation of multiple indicators may be found, for example, in chapters 7 and 8 of Blalock (1974).

IV. LONGITUDINAL PANEL DESIGNS

One of the most common and useful designs for longitudinal research is the longitudinal panel design where the same sample of units, is observed at more than one point in time. Essentially, panel designs are a combination of time-series and cross-sectional designs, with a cross-section (wave) being measured at each time point. Typically, the number of cases in each wave is considerably greater than the number of waves. Longitudinal panel designs are often employed to study reciprocal causation, using "longitudinal (or intertemporal) variation to disentangle reciprocal causal effects" (Hannan & Young, 1974, p. 2). Hannan & Young presented another use for panel designs:

> A closely related motivation for panel analysis arises from work with models containing unobservable variables. Such models confront measurement and other analytic difficulties by inserting into structural equations models both measured and unmeasured variables. The use of unobservables will ordinarily lead to problems of identification unless strong restrictions are placed on the model. One possibility that occurred to a number of sociologists is to measure the same variables at multiple points in time and presume that the causal relations under study are time-invariant. Under a limited number of conditions this strategy leads to identification of multi-variable, multi-wave panel models containing unobservables. The main point for present purposes is that this use of the panel design uses temporal variation to eliminate identification problems. (p. 5)

The assumption that the causal relations under study are time invariant means that the configurations of the constructs are assumed identical across the different time periods. This is equivalent to an assumption of "explanatory continuity". "Explanatory discontinuity occurs when behavioral antecedents or mediating processes at one period of development differ from those at another period" (Huston-Stein & Baltes, in press). In a causal model of a longitudinal panel, explanatory discontinuity would mean that the causal orderings differ among the waves, and a model assuming invariance over time would be misspecified in some waves. No assumptions that the causal parameters are invariant over time is being made; the parameters are allowed to vary.

A. Two-Wave Panel Designs

The simplest (although not the ideal) panel design is the two wave panel, with observations recorded at two points in time. In the most popular and widely discussed two-wave design, two variables are measured at each time point and thus the design is termed the two-wave-two-variable panel (2W2V). The representation of the causal model usually assumed for 2W2V panels is shown in Figure 4. Some important restrictions are built into this model to allow estimation of the parameters. Most important is the assumption that lagged causation is the sole causal force; simultaneous causation between X_2 and Y_2 is ruled out. Also, the residuals (disturbance terms) u and v are assumed to be independent, and the variables must be measured without error. More general models are considered by Duncan (1969, 1972) who concluded:

> no set of 2W2V data will answer a question about direction of causal influence or relative importance of causes except on some set of definite assumptions. If one wishes to avoid assumptions of the type illustrated here, the only recourse is to expand the study design beyong the limits of 2W2V (1969, p. 181).

Insert Figure 4 about here

The configuration in Figure 4 can be represented as a regression model:

$$X_2 = \beta_1 X_1 + \gamma_2 Y_1 + u$$

$$Y_2 = \beta_2 X_1 + \gamma_1 Y_1 + v$$

The parameters, β_1, β_2, γ_1 and γ_2 are estimable by regression, and the estimated causal effects are easily interpreted, when the restrictions of the model are satisfied.

A research example of the use of the 2W2V design is the investigation of Crano, Kenny, and Campbell (1972) concerning the causal relationship between intelligence and achievement. The substantive background for the study was that:

> The literature of cognitive development has produced two opposing models of mental growth. One holds that the acquisition of concrete

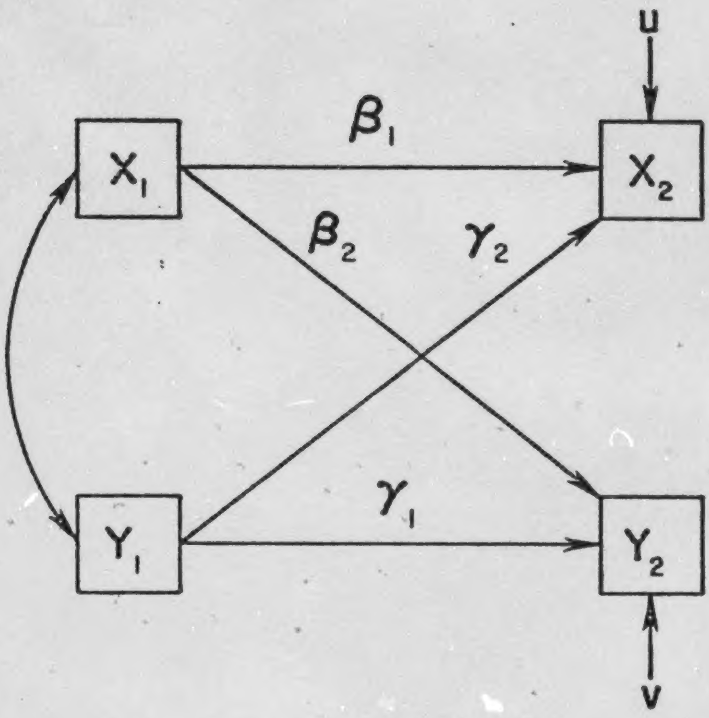

u
β_1
X_1
X_2
β_2
γ_2
γ_1
Y_1
Y_2
v

Figure 4 -- Representation of the causal model for 2W2V panel. Assumptions built into this model are no simultaneous causation, measurements without error and uncorrelated residuals.

> mental skills causes the later development of higher order organizational schema or rules. The contrasting model postulates a progression in which the initial acquisition of larger schema results in the increased capacity to acquire new concrete skills. While both probably operate to some extent, an attempt was made in this research to determine the preponderant developmental sequence. (p. 258).

The research question was:

> Does the acquisition of specific skills or the learning of specific information (Achievement) result in an increased ability for abstraction (intelligence), or is the progression more accurately described as one in which intelligence causes achievement, that is, does the greater ability to form abstractions result in a greater amount of concrete information being absorbed and retained? (p. 258-259)

The research design consisted of measurements of intelligence and achievement at two points in time. The data are intelligence (Lorge-Thorndike) and achievement (Iowa Basic Skills) test scores of children attending the fourth grade in the academic year 1963-4 and the same tests given two years later to the same students. If the required assumptions were made, the data can be entered into the model of Figure 4 (with X as Intelligence and Y as Achievement), and the causal parameters, estimated. Crano et al. did not perform a causal model analysis of their 2W2V data, but instead employed the method of cross-lagged correlations discussed below.

In the education and psychological literature, much attention has been given to cross-lagged correlations as a means for inferring the direction of causal influence. In Figure 4 the cross-lagged correlations are $r_{X_1Y_2}$ and $r_{Y_1X_2}$. If $r_{X_1Y_2} > r_{Y_1X_2}$ the suggested interpretation (Campbell, 1963) is that X causes Y. Although the cross-lagged procedure has come under sharp attack (Duncan 1969, Goldberger 1971) its use is still widely recommended (e.g., Clarke-Stewart 1973; 1975; Huston-Stein & Baltes, in press). Goldberger charged that the justification of the cross-lagged correlation method is based not on evidence but on "intuition, hunch, and plausibility, which are hardly persuasive" (1971, p. 102). A superior method for inferring the direction of causal influence is to estimate the parameters of the causal model (Figure 4) which have direct interpretation.

The cross-lagged correlations are a complicated function of the causal parameters β_1, β_2, γ_1, γ_2. It is easy to construct cases where cross-lagged correlations give exactly the wrong answer; that is lead to the inference that the direction of causality is opposite to that in the underlying model. For example in the causal model (Figure 4) let all variables have unit variance and let the true values of the causal parameters be $\beta_1 = 0.9$, $\beta_2 = 0.3$, $\gamma_1 = 0.1$, $\gamma_2 = 0$, and $r_{X_1Y_1} = 0.6$. Thus the underlying causal influence over time is in the direction of X causes Y. The cross-lagged correlations are $r_{X_1Y_2} = .36$ and $r_{Y_1X_2} = .54$, and the conclusion would be that Y causes X. This example is not just a mathematical aberration; large differences in stability of the two measures are common, for example, in research on mother-child interactions (see Clarke-Stewart, 1973) and will cause the cross-lagged analysis to break down. Estimation of the underlying causal model is preferred to developing a special theory of cross-lagged correlations.

B. Complications in 2W2V model

As Duncan (1972) sagely observed, it is unreasonable to expect "that in panel analysis the usual obstacles to inference and estimation are suspended for the benefit of the analyst" (p. 37). Two major obstacles are measurement error and specification error. Their consequences for analysis are discussed below in the context of the 2W2V design. The consequences of violation of assumption and the consideration of alternative models is applicable to all causal models.

A 2W2V model with measurement error is shown in Figure 5. The causal influences are presumed to be transmitted through the true scores X*, Y* which are unmeasured. Because additional parameters must be included in the model to represent the measurement error there is no longer enough information from the observables to estimate the causal paths, and the model is underidentified. The problems with cross-lagged correlations are naturally increased when the variables are measured with error. Wiley and Hornik (1973) argued cogently against their use:

> Comparison of cross-lagged panel correlations was the first widely advocated non-experimental technique for the attribution of causality to quantitatively scaled variables in social-psychological research. In addition to the lack of a clearly stated statistical model, one universally recognized weakness of this technique is its serious distortion by commonly occurring measurement errors. One especially distorting event is systematic change in the reliability of variables over time. Such events occur, for example, when a true variable's variance changes, although the quality of the measurement remains the same. This is nearly always the case when any change in a true variable takes place (Wiley & Wiley, 1970). Therefore, it is difficult to justify the use of [this] technique in typical social research settings. (p.2)

Insert Figure 5 about here

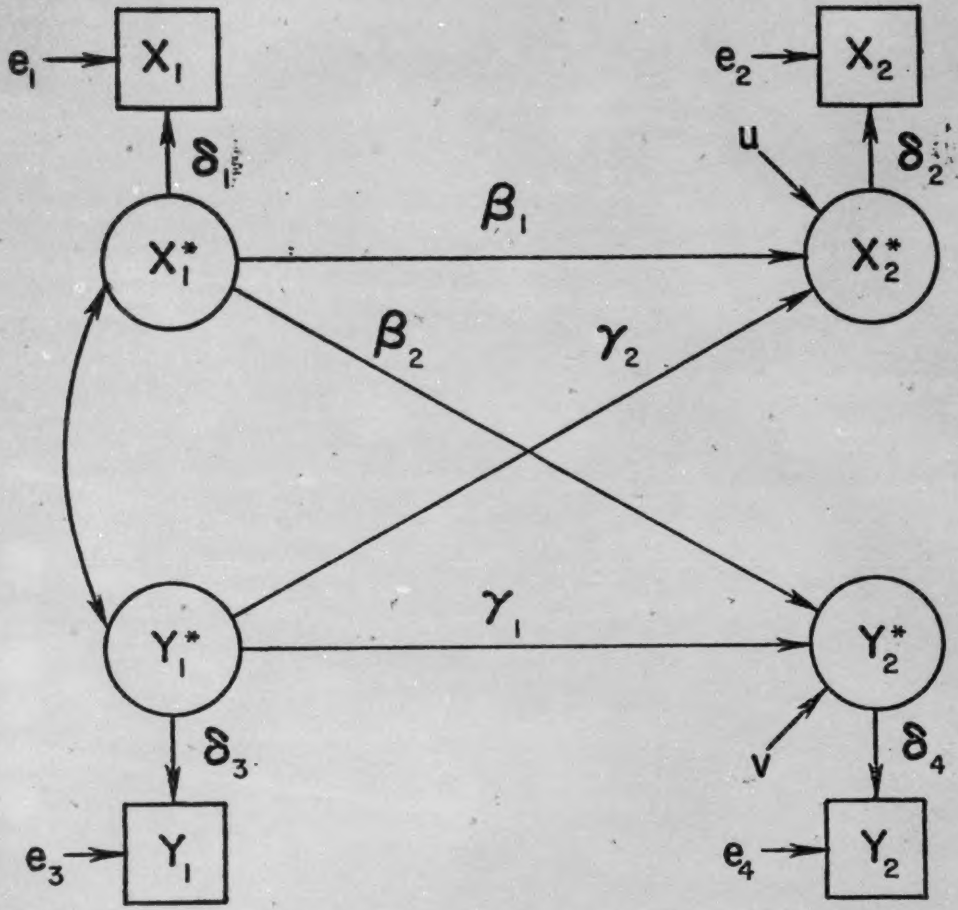

e_1
X_1
e_2
X_2
δ_1
u
δ_2
β_1
X_1^*
X_2^*
β_2
γ_2
γ_1
Y_1^*
Y_2^*
δ_3
v
δ_4
e_3
Y_1
e_4
Y_2

Figure 5 -- Representation of a 2W2V design with measurement error. In addition to the previous assumptions for 2W2V panels the measurement errors (e_i) are assumed uncorrelated. However, the causal parameters of interest β_1, β_2, γ_1, γ_2, which represent the causal influences between the unmeasured true scores, cannot be estimated without additional information.

If independent estimates of the reliabilities are not available, the additional information needed to estimate the model may be obtained by two strategies: increasing the number of measures of each variable at each point (Wiley & Hornik, 1973) or increasing the number of time points to three (Heise, 1969). Wiley and Hornik (1973) explain the use of duplicate measures represented Figure 6:

> We suggest expanding the number of measures of each variable at each point. Two measures of each variable at each time point produce enough additional information to allow the calculation of all the quantities in a quite general two-time point model. (p. 8).

Insert Figure 6 about here

Specification error occurs when the postulated causal model is an incorrect representation of the actual (psychological) process. A commonly considered form of specification error in panel analysis is the existence of a common factor causing both the X and Y variables (Duncan 1972; Kenny, 1973). Duncan presents many examples of specification errors due to omitted causal influences and correlated errors. Clearly, numerous ways exist for assumptions to be violated. One example presented by Duncan (1972, p. 56) is adapted as Figure 7. The data reported by Jöreskog (1973a, see also Chapter this volume) are test results in mathematics and science taken in the fifth and seventh grade. It may be reasonable to posit that a common factor (F) is the sole causal agent; the observed test scores are related only through the common cause. An analysis based on the model in Figure 4 when the model in Figure 7 is the correct one will yield misleading results. Many other plausible forms of specification error in longitudinal panel models are considered in Duncan (1972, 1975a) and Jöreskog and Sörbom (1976a).

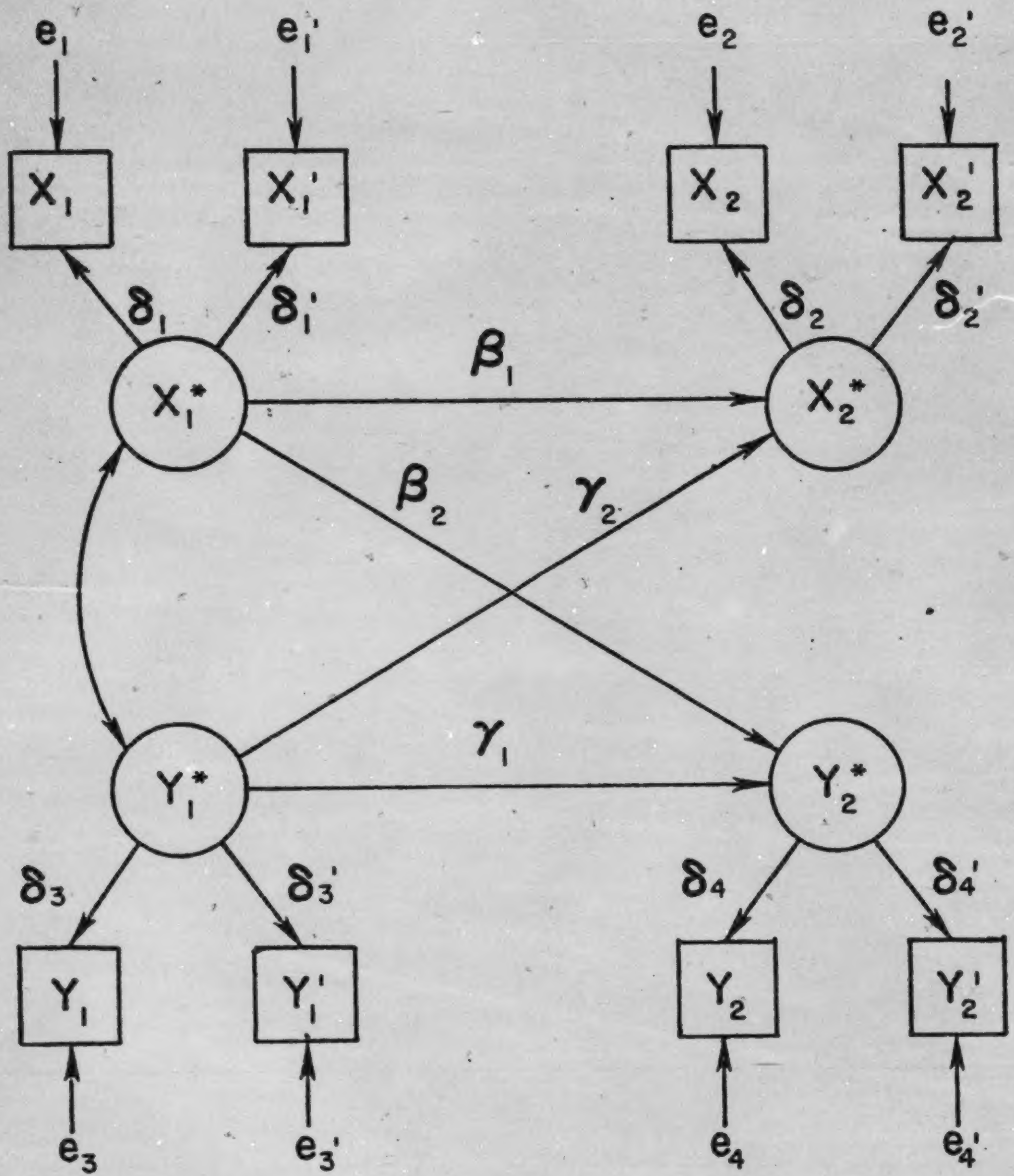

e1
e1'
e2
e2'
X1
X1'
X2
X2'
δ1
δ1'
δ2
δ2'
β1
X1*
X2*
β2
γ2
γ1
Y1*
Y2*
δ3
δ3'
δ4
δ4'
Y1
Y1'
Y2
Y2'
e3
e3'
e4
e4'

Figure 6 -- A 2 wave 2 variable 2 indicator (2W2V2I) model. The duplicate measures of each causal variable allow estimation of the causal parameters, β_1, β_2, γ_1, γ_2.

Insert Figure 7 about here

C. Multiple Wave Panel Designs

In developmental research, more than two waves of observations are usually desired on substantive grounds. Multiple waves have the added advantage that they allow estimation of parameters under less restrictive assumptions (Hannan, Rubinson & Warren, 1974). However, as might be expected, multiple waves of observations make model specification and estimation more complex. To illustrate the use of panel designs, two studies in which the investigators attempted to make causal statements from longitudinal data are formulated as causal models. The original analyses are discussed so as to point out the advantages of causal models in design and analysis.

Clarke-Stewart (1973) collected data on 36 mother-child dyads over a nine-month period, tracing the children from age nine months to 18 months. Repeated observations of the mother-child dyads were made both in natural settings (the home) and in a variety of standardized or semistructured (test and laboratory) situations. Clarke-Stewart began her longitudinal study by identifying as one of the problems of method in child development research that "in the past, studies have seldom been longitudinal in design, and when longitudinal, maternal behavior has often been measured at one time and infant performance at another, consequently not permitting the analysis of causal relationships" (p. 8). Since the mother-child interaction is certainly an instance of reciprocal causation, longitudinal designs may allow the untangling of the reciprocal causal influences.

The portion of the total data which Clarke-Stewart used to make causal inferences (see 1973, pp. 82-91) corresponds to a three wave longitudinal panel with waves at the child ages of 11, 14 and 17 months. Most of the attempts at

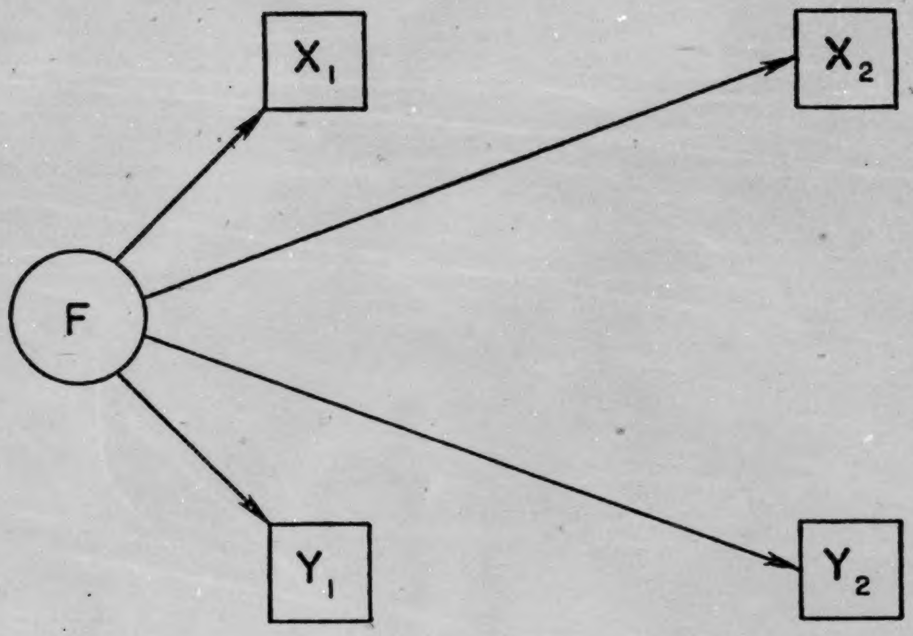


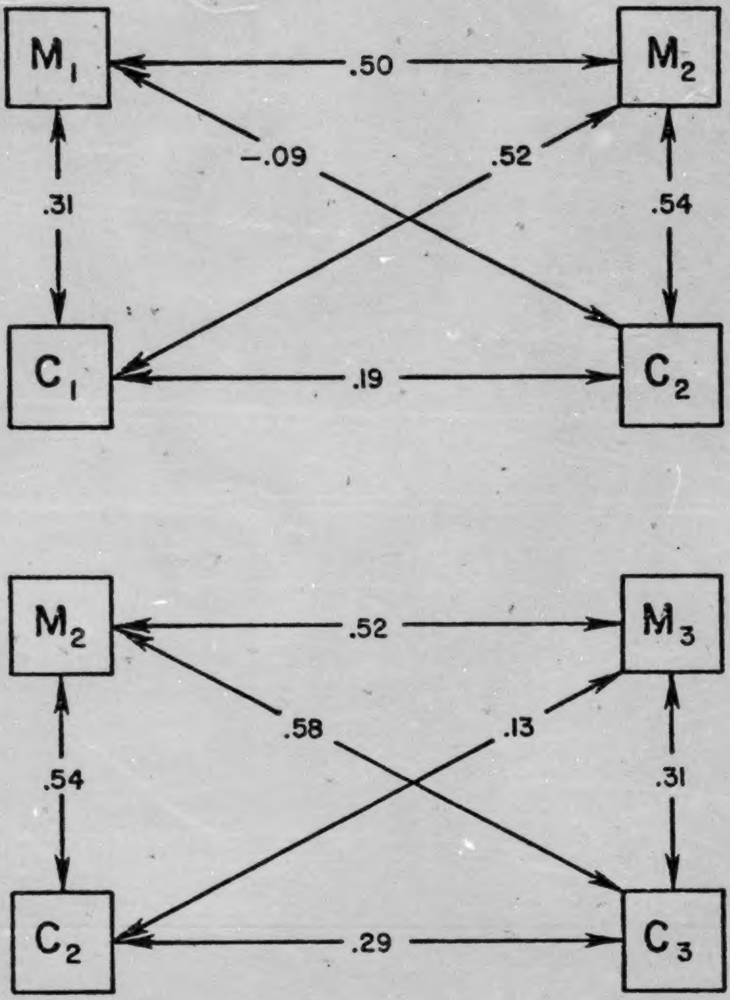
Figure 7 -- An alternative causal structure for the 2W2V design. All the causal influence is carried by the common factor F.

causal explanation used data from only the first and third waves. For all causal statements Clarke-Stewart used the method of cross-lagged correlations. One instance in which she considered all three waves is illustrated in Figure 8 where the causal influences of maternal attention and the child's attachment to the mother are examined. As can be seen in Figure 8, Clarke-Stewart considered the three waves of data as two two-wave pieces. The somewhat ambiguous causal conclusion drawn from the two sets of cross-lagged correlations is

> At Times 1 and 2 the cross-lagged correlations for infant attachment and maternal attention suggested that maternal attention was causing an increase in infant attachment. From Time 2 to Time 3 however, the cross-lagged correlations implied the opposite: that infant attachment was causing maternal attention. This finding suggests the possibility that, as mother and child search for harmonious, balanced interaction over the course of development, first one then the other assumes the "causal role". (p. 91)

Insert Figure 8 about here

A causal model of the 3W2V panel is shown in Figure 9. The panel model in Figure 9 allows doubly lagged causation between waves one and three between variables but not within variables. Labouvie (1974) would term this model a between-domain distal-within domain proximal model. A causal model analysis of these three-wave data is certainly more attractive since it would use all the data at once as opposed to breaking the longitudinal data into two separate two wave sequences. Also it might be expected that analyzing all three time periods at once would alter the original conclusions, especially if there were


M_1
.50
M_2
−.09
.52
.31
.54
C_1
.19
C_2
M_2
.52
M_3
.58
.13
.54
.31
C_2
.29
C_3

M_T = Maternal Attention to Child at Time T

C_T = Child's Attachment to Mother at Time T

Figure 8 -- The 2W2V configuration, with the cross-lagged and synchronous correlation exhibited, of a three wave design. (Adapted from Clarke-Stewart, 1973).

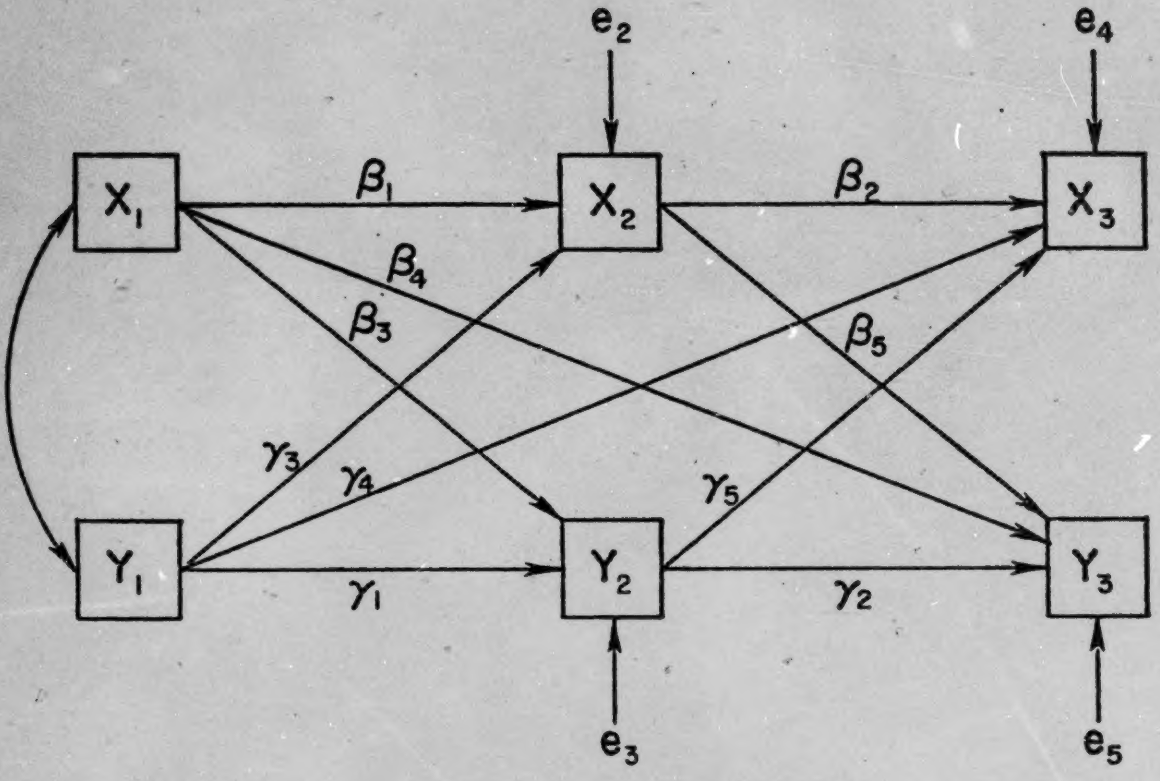
any appreciable direct causal effect between wave one and wave three, represented by β_4 and γ_4 in Figure 9.

Insert Figure 9 about here

O'Mally and Bachman (1976) reported an eight year longitudinal study in which the relationship between self-esteem and success was investigated. Measures of self-esteem were collected at five points in time on 1600 young men over an eight year period (1966-1974) beginning when the subjects entered tenth grade through five years after graduation. Success was measured once, by level of occupational and educational attainment in 1974. Many other variables were measured across some or all of the five waves: status of aspired occupation at all waves, grades at the first three waves, and background measures of ability and SES at the first wave. The design was not strictly a longitudinal panel since the major outcome variable, success, was only measured at the final wave. Therefore, the reciprocal relationship between self-esteem and success (attainment) may be difficult to unravel.

O'Mally and Bachman distinguish three logically independent patterns of causation between self-esteem and attainment: (a) that self-esteem contributes directly to attainment; (b) that attainment contributes, both indirectly and directly, to heightened self-esteem; and (c) that some of the underlying determinants of self-esteem are also important determinants of attainment (e.g., academic ability, past educational accomplishments and family socioeconomic level).

These three patterns of causation, which may be responsible for the observed relations between self-esteem and success, can be represented:


e_2
e_4
X_1
β_1
X_2
β_2
X_3
β_4
β_3
β_5
γ_3
γ_4
γ_5
Y_1
Y_2
Y_3
γ_1
γ_2
e_3
e_5

$$
\begin{aligned}
X_2 &= \beta_1 X_1 + \gamma_3 Y_1 + e_2 \\
X_3 &= \beta_2 X_2 + \gamma_4 Y_1 + \gamma_5 Y_2 + e_4 \\
Y_2 &= \gamma_1 Y_1 + \beta_3 X_1 + e_3 \\
Y_3 &= \gamma_2 Y_2 + \beta_4 X_1 + \beta_5 X_2 + e_5
\end{aligned}
$$

Figure 9 -- The 3W2V design of Clarke-Stewart (1973) with the causal parameters and structural regression equations shown. The X variables are Maternal Attention to the Child and the Y variables are the Child's Attachment to the Mother.

(a) Self-esteem → Attainment

(b) Attainment → Self-esteem

(c)
$$\begin{tikzcd} & \text{Attainment} \\ \text{Background Variables} \arrow[ur] \arrow[dr] & \\ & \text{Self-esteem} \end{tikzcd}$$

The analysis of these causal sequences is based entirely on correlations. To illustrate some of the problems with the analysis, consider the correlational analysis O'Mally and Bachman (1976) proposed for the second suggested causal sequence, that attainment (their B) causes self-esteem (A). Two of their pertinent hypothses were:

> Hypothesis 2a. Later self-esteem is positively correlated with attainment, after statistically controlling background and ability and also earlier self-esteem (based only on B causes A).
>
> Hypothesis 3. Attainment (five years after high school) is more strongly correlated with later self-esteem than with earlier self-esteem. This hypothesis is based on B causes A. (p.8)

Figure 10 represents a causal model of the relationship of self-esteem and attainment, relevant to the above hypotheses.

Insert Figure 10 about here

Figure 10 shows that the O'Mally and Bachman hypothesis that B causes A is equivalent to postulating a nonzero value of γ. However, a positive value of β_5 would cause the correlation between self-esteem at time point 5 and attainment, with earlier self-esteem held constant, to be positive even when γ is zero. Therefore Hypothesis 2a is not "based only on B causes A." By the

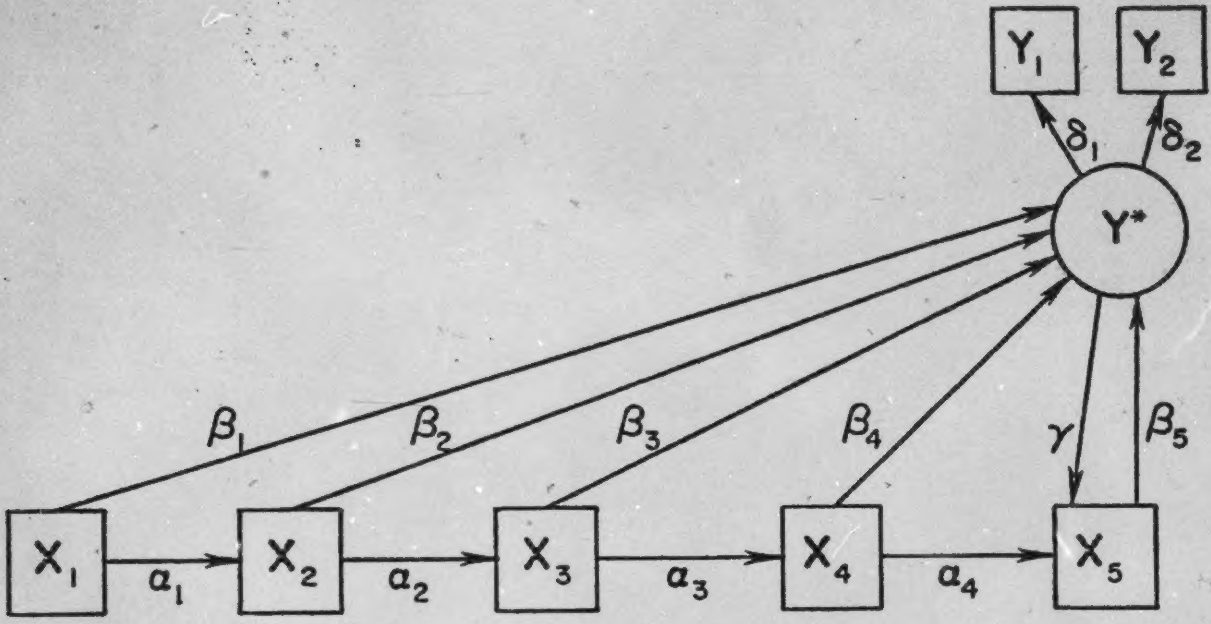

Y₁
Y₂
δ₁
δ₂
Y*
β₁
β₂
β₃
β₄
γ
β₅
X₁
a₁
X₂
a₂
X₃
a₃
X₄
a₄
X₅

Figure 10 -- A causal model of esteem and attainment patterned after the O'Mally and Bachman (1976) data. X_1 through X_5 are self-esteem at the five time points. Y* is the attainment (success) construct with indicators Y = educational attainment, Y_2 = occupational attainment. The causal paths and parameters of the causal relations are shown.

same logic, if β_1, β_2, β_3, β_4 were zero in Figure 10, then attainment would be more strongly correlated with later self-esteem (X_5) than with earlier self-esteem for any nonzero values of γ and β_5. Hypothesis 3, then, is also not properly tested when correlations are used in place of estimation of γ. The correlations confound the effects of a nonzero γ with other causal configurations. Because plausible alternative explanations exist for the correlations that O'Mally and Bachman posit as proof of their causal hypotheses, Hypotheses 2a and 3 are not logically sound representations of the causal structure.

In addition to the previously discussed problems with correlations as a measure of causation, there are important ones associated with standardization. Wiley & Wiley (1970) and Tukey (1954) among others warned against using standardized measures for causal statements. It is unreasonable to expect variances to remain constant over time; in fact the increase over time of interindividual differences in development implies an increase in variance. But a change in the variance of, say, self-esteem changes the correlation without necessarily affecting the strength of the relationship.

Furthermore, aside from the inadequacies of correlations in this context, without repeated measures of attainment over the waves of observations, O'Mally and Bachman are not able to disentangle unambiguously the reciprocal causation between self-esteem and attainment. Perhaps grades could have been used as a proxy for attainment in the earlier waves, in order to evaluate causal influences from the available data.

D. Distal Causation and Causal Lag

The question of the proper degree of causal lag has frequently appeared in the causal models that have been considered. For example, in the 2W2V models simultaneous causation was ruled out; consequently, the proper causal lag was assumed to be one time period (see Figure 4). In multiwave models such as Figure 9 (Clarke-Stewart's 3W2V design), a variety of causal lag patterns are possible. There may be direct effects from time 1 to time 3 both within and between the variables, in addition to simultaneous effects. In the O'Mally and Bachman example, a number of different indirect and direct causal lag patterns are possible.

A major concern of life-span psychology is establishing long-term causal relations. "Since life-span research deals most explicitly with the analysis of long-term phenomena and chains, it has been forced to attend to the issue of structuring time-lagged relationships and performing distal-cause analyses (Huston-Stein & Baltes, in press, p. 11). Often the results of the causal model analysis of the time-lagged relationships depend on the kinds of causal lags built into the model. Unfortunately, developmental theories which would be the best guide to the proper kinds of causal lags to incorporate in the model offer little guidance in most situations. The majority of longitudinal panel models postulate single period causal lags or equal lags across multiple waves. An exception is an analysis by Hannan et al. (1974) whose 3W2V model has a two-wave causal lag in one direction, and a one-wave lag in the other direction.

V. SUMMARY

This chapter has involved consideration of some methods for making causal statements in developmental research. In the study of development and growth, longitudinal designs should be used. Cross-sectional designs will give conflicting (and misleading) evidence over replications, and can yield little information on the causal dynamics of development. From practical necessity and for ecological validity the longitudinal designs will often be nonexperimental and naturalistic. Additional knowledge from substantive theory and empirical evidence may be profitably incorporated into the data analysis of such designs through the use of causal models.

Causal models are presented as an attractive vehicle for the formulation of well-specified theory in an empirically testable form. The construction of causal models is a two-stage procedure involving the structural model which specifies the relations between the theoretically important constructs and the measurement model which relates unobserved constructs to their observable measures. The construction of a causal model from a substantive study illustrates the usefulness of multiple measures of constructs in the estimation of causal effects.

Longitudinal panel designs combine features of cross-sectional and time series designs in that waves of cross-sections, containing the same sampling units, are measured at two or more time points. A strong feature of panel designs is their ability, in certain situations, to separate reciprocal causal influences between two or more variables, which is in general impossible with cross-sectional data. The two-wave two-variable (2W2V) design has been widely used for this purpose. A formulation of the 2W2V design as a causal model reveals that highly restrictive assumptions are necessary for estimation of causal influences and for the desired determination of which variable causes the other. The popular method of cross-lagged correlations is shown to be undesirable; direct estimation of the relevant causal parameters is always preferable.

Two major technical problems associated with causal models (and with almost all analysis techniques) are measurement error and specification error. However, multiple indicators of the fallibly measured constructs may, in many instances, allow unbiased estimation of the model. Specification errors arise most often when important causal variables are omitted from the model and cause the estimates of the causal influences between the included parameters to be biased. Some common forms of these two obstacles to estimation are illustrated for the 2W2V design. Technical issues in estimation are treated in another chapter.

The more powerful multiwave panel designs are described by formulating causal models from two substantive longitudinal studies. The use of correlations in these studies to support causal statements is criticized.

Causal models specify on an a priori basis how the same correlations might have arisen from a variety of causal mechanisms. The value of causal models, then, lies in their incorporation of substantive theory and reduced dependence on post hoc interpretation of correlational data.

STATISTICAL ESTIMATION OF STRUCTURAL MODELS IN LONGITUDINAL-DEVELOPMENTAL INVESTIGATIONS

Karl G. Jöreskog

University of Uppsala
Uppsala, Sweden

Table of Contents

STATISTICAL ESTIMATION OF STRUCTURAL MODELS IN LONGITUDINAL-DEVELOPMENTAL INVESTIGATIONS

ABSTRACT
I. INTRODUCTION
II. COMPUTER PROGRAMS
III. ESTIMATION OF GROWTH CURVES
IV. TWO-WAVE MODELS
V. MULTI-WAVE MODELS
VI. SUMMARY

ABSTRACT

This chapter deals with statistical methodological problems in the analysis of data from large longitudinal studies where the same or similar quantitative measurements have been obtained at two or more occasions, possibly from several different groups of people. Several models are developed for a wide range of applications for psychological and educational measurements. The problems of model specification, statistical identification, estimation and testing are discussed. In particular, the chapter focuses on the following problems: (i) the estimation of growth curves under auto-regressive models; (ii) the treatment of measurement errors in observed variables; and (iii) the scaling of latent variables. Several examples are given illustrating the assessment of fit of a model and data-analytic strategies for model modification.

I. Introduction

The characteristic feature of a longitudinal research design is that the same measurements are obtained from the same people at two or more occasions. The purpose of a longitudinal or panel study is to assess the changes that occur between the occasions and to attribute these changes to certain background characteristics and events existing or occurring before the first occasion and/or to various treatments and developments that occur after the first occasion. A schematic illustration of a two-wave longitudinal design is given in Fig. 1. Earlier chapters of this volume have dealt with the conceptual and substantive issues and with the logic of causal model building in longitudinal research in developmental psychology and education. Other chapters have dealt with specific methodological problems. Wiley and Harnischfeger (1973) have given an account of the conceptual issues in the attribution of change in educational studies. In the sociological literature there has been a number of articles concerned with the specification of models incorporating causation and measurement errors and the analysis of data from panel studies (see e.g. Bohrnstedt, 1969; Heise, 1969, 1970; Duncan, 1969, 1972, 1975b). Other papers dealing with methodological problems are Lord (1963), Thorndike (1966), Härnqvist (1968), Cronbach and Furby (1970) and Bergman (1971). Complex models involving multiple measurements and/or several occasions have been considered by Harris (1963), Jöreskog (1970a), Corballis and Traub (1970), Nesselroade (1972), Corballis (1973), Bentler (1973), Frederiksen (1974), Jöreskog and Sörbom (1976a-b) and Olsson and Bergman (1977).

Insert Figure 1 about here

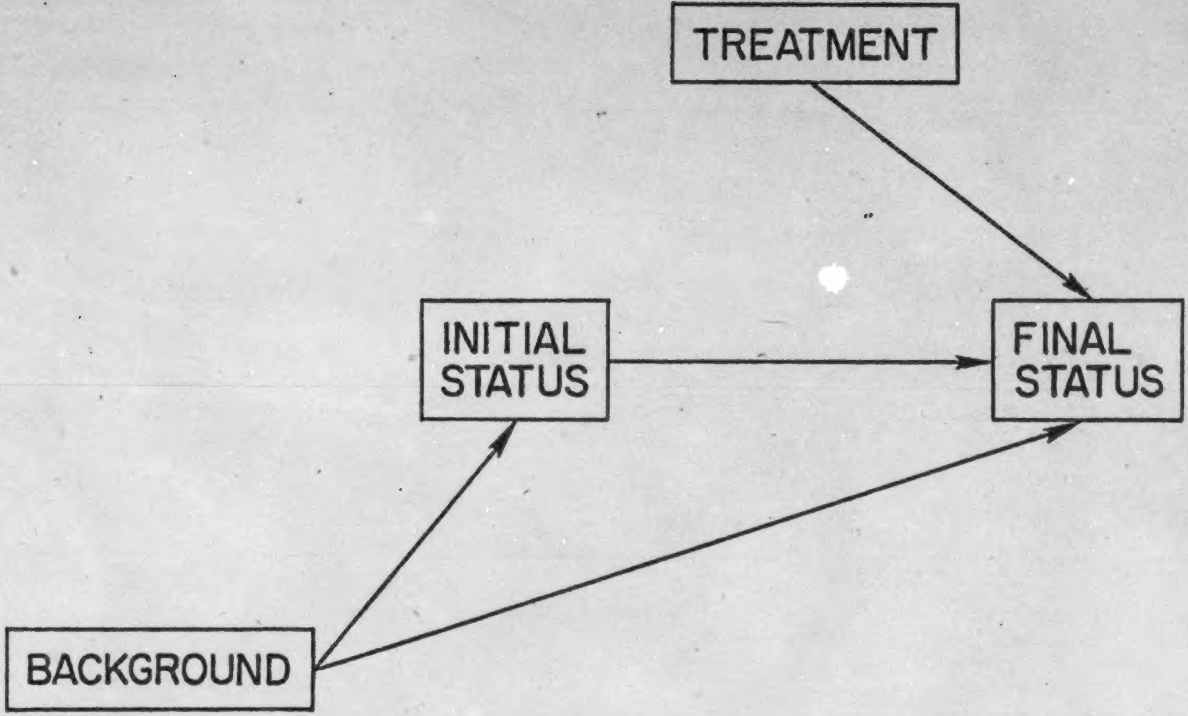

TREATMENT
INITIAL STATUS
FINAL STATUS
BACKGROUND

Figure 1 -- Schematic representation of a two-wave longitudinal design.

In this paper we consider several models suitable for analyzing longitudinal data and deal with problems of model specification and statistical identification, estimation and testing. The general setup is that of a longitudinal study where the same or similar quantitative measurements have been obtained at two or more occasions, possibly from several different groups of people. The models cover a wide range of applications and are relevant for psychological and educational measurements as well as for social and socio-economic measurements.

Section III considers the measurement and assessment of change at the group level. It deals with the estimation of growth curves describing the means of the variables as functions of time. This section also considers various auto-regressive or first-order Markov models that occur naturally in repeated measurements. In this section we treat the variables as errorfree.

One of the most difficult problems for a social scientist, when it comes to the formulation of a causal model, arises because many of the concepts and constructs that he/she wants to work with are not directly measurable (see e.g. Torgerson, 1958, Chapter 1, Goldberger, 1972, Duncan 1975a, and Heise 1975). Although such hypothetical concepts and constructs, or latent variables, as we shall call them, cannot be directly measured, a number of variables can be used to measure various aspects of these latent variables more or less accurately. Thus, while the latent variables cannot be directly observed, they have operational implications for relationships among observed variables. We may regard the observed variables as indicators of the latent variables. Each indicator has a relationship with the latent variable, but if we take one indicator alone to measure the latent variable we would obtain a biased measurement. By using several indicators of each latent variable we get a better measurement of the latent variable .

Another reason for using latent variables in behavioral and socioeconomic studies is that most of the measurements employed contain sizeable errors of measurement (observational errors) which, if not taken into account, can cause severe bias in the results. Errors of measurement arise because of imperfection in the various measurement instruments (questionnaires, tests, etc.), that are used to measure such abstractions as people's behavior, attitudes, feelings and motivations. Even if we could construct valid measurements of these it is usually impossible to obtain perfectly reliable variables. Special care must be taken to obtain measurements that really measure the latent traits or hypothetical constructs that one is interested in measuring. Various models with latent variables are considered in sections IV for two-wave situations and in section V for multi-wave situations.

A common experience in two-wave longitudinal studies is that the initial status is the best determinant or predictor of the final status (see e.g. Lord, 1963). Therefore, if one is interested in attributing change to certain background variables one must find some way of effectively eliminating the initial status from the final status. This has been taken to mean that one should study difference scores (final scores minus initial scores). However, this is not necessary; the important thing is that both background variables and initial measures are included in the model as determinants of final measures. In multi-wave studies one can determine the effect of the background variables on the dependent variable at various points in time. Most of the models introduced in sections IV and V are considered both with and without background variables.

Often it is not possible, or even desirable, to specify the model completely since there may be other models which are equally plausible. In such a situation it is necessary to have a technique of analysis which will give information about which of a number of alternative models is (are) the most reasonable. Also, if there is sufficient evidence to reject a given model due to poor fit to the data, the technique should be such as to suggest which part of the model is causing the poor fit. Several examples will be given illustrating the assessment of fit of a model and strategies for model modification.

In presenting the various models it is convenient to use a path diagram. In this path diagram observed variables are enclosed in squares whereas latent variables are enclosed in circles. Residuals (errors in equations) and errors of measurements are included in the diagram but are not enclosed. A one-way arrow pointing from one variable x to another variable y indicates a possible direct causal influence of x on y, whereas a curved two-way arrow between x and y indicates that x and y may correlate without any causal interpretation of this correlation being given. It is convenient to write the coefficient associated with each arrow in the path diagram. When the coefficient is omitted it means that it is one. For one-way arrows such coefficients will be (partial) regression coefficients (path coefficients) and for two-way arrows they will be covariances. In the special case when all observed and latent variables are standardized, these coefficients will be correlations. With these conventions it is possible to write down the model equations from the path diagram. In order to define the model completely it is only necessary to specify the assumptions about the origin and unit of measurement of the variables involved and the distributional assumptions, if any.

II. Computer Programs

All the models considered in this chapter may be estimated by means of three computer programs: ACOVSM by Jöreskog, van Thillo and Gruvaeus (1970), LISREL by Jöreskog and Sörbom (1976c) and COFAMM by Sörbom and Jöreskog (1976). The general models on which these programs are based are described briefly here for future reference. For each model introduced in later sections of this chapter it will be shown how this is a special case of one of the models presented here.

A. ACOVSM

The ACOVSM model considers a data matrix $\underset{\sim}{X}(N \times p)$ of N observations on p variables and assumes that the rows of $\underset{\sim}{X}$ are independently distributed, each having a multivariate normal distribution with the same variance-covariance matrix $\underset{\sim}{\Sigma}$. It is assumed that

$$E(\underset{\sim}{X}) = \underset{\sim}{A}\underset{\sim}{\Xi}\underset{\sim}{P}, \tag{1}$$

where $\underset{\sim}{A}(N \times g) = (a_{\alpha s})$ and $\underset{\sim}{P}(h \times p) = (p_{ti})$ are known matrices of ranks g and h, respectively, $g \leq N$, $h \leq p$, and $\underset{\sim}{\Xi}(g \times h) = (\xi_{st})$ is a matrix of parameters; and that $\underset{\sim}{\Sigma}$ has the form

$$\underset{\sim}{\Sigma} = \underset{\sim}{B}(\underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^2)\underset{\sim}{B}' + \underset{\sim}{\Theta}^2, \tag{2}$$

where the matrices $\underset{\sim}{B}(p \times q) = (\beta_{ik})$, $\underset{\sim}{\Lambda}(q \times r) = (\lambda_{km})$, the symmetric matrix $\Phi(r \times r) = (\phi_{mn})$, and the diagonal matrices $\Psi(q \times q) = (\delta_{kl}\psi_k)$ and $\Theta(p \times p) = (\delta_{ij}\theta_i)$ are parameter matrices. δ_{ij} denotes the Kronecker delta, which is one if $i = j$ and zero otherwise.

Thus the general model is one where means, variances, and covariances are structured in terms of other sets of parameters that are to be estimated. In any application of this model, p, N, and X will be given by the data, and g, h, q, r, A, and P will be given by the particular application. In the special case when both Ξ and Σ are unconstrained, one may test a sequence of linear hypotheses of the form

$$C \Xi D = 0 , \tag{3}$$

where $C(s \times g)$ and $D(h \times t)$ are given matrices of ranks s and t, respectively.

For further information about the ACOVSM model and its uses see Jöreskog (1970c, 1973b).

B. LISREL

The LISREL model considers random vectors $\eta' = (\eta_1, \eta_2, \ldots, \eta_m)$ and $\xi' = (\xi_1, \xi_2, \ldots, \xi_n)$ of latent dependent and independent variables, respectively, and the following system of linear structural relations

$$B\eta = \Gamma\xi + \zeta \tag{4}$$

where $B(m \times m)$ and $\Gamma(m \times n)$ are coefficient matrices and

$\underset{\sim}{\zeta}' = (\zeta_1, \zeta_2, \ldots, \zeta_m)$ is a random vector of residuals (errors in equations, random disturbance terms). Without loss of generality it may be assumed that $E(\underset{\sim}{\eta}) = E(\underset{\sim}{\zeta}) = \underset{\sim}{0}$ and $E(\underset{\sim}{\xi}) = \underset{\sim}{0}$. It is furthermore assumed that $\underset{\sim}{\zeta}$ is uncorrelated with $\underset{\sim}{\xi}$ and that $\underset{\sim}{B}$ is nonsingular.

The vectors $\underset{\sim}{\eta}$ and $\underset{\sim}{\xi}$ are not observed but instead vectors $\underset{\sim}{y}' = (y_1, y_2, \ldots y_p)$ and $\underset{\sim}{x}' = (x_1, x_2, \ldots x_q)$ are observed, such that

$$\underset{\sim}{y} = \underset{\sim}{\Lambda}_y \underset{\sim}{\eta} + \underset{\sim}{\varepsilon}, \tag{5}$$

$$\underset{\sim}{x} = \underset{\sim}{\Lambda}_x \underset{\sim}{\xi} + \underset{\sim}{\delta}, \tag{6}$$

where $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$ are vectors of errors of measurement in $\underset{\sim}{y}$ and $\underset{\sim}{x}$, respectively. $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are assumed to be measured as deviations from their means. The matrices $\underset{\sim}{\Lambda}_y(p \times m)$ and $\underset{\sim}{\Lambda}_x(q \times n)$ are regression matrices of $\underset{\sim}{y}$ on $\underset{\sim}{\eta}$ and of $\underset{\sim}{x}$ on $\underset{\sim}{\xi}$, respectively. It is convenient to refer to $\underset{\sim}{y}$ and $\underset{\sim}{x}$ as the observed variables and $\underset{\sim}{\eta}$ and $\underset{\sim}{\xi}$ as the latent variables. The errors of measurement are assumed to be uncorrelated with the latent variables.

Let $\underset{\sim}{\Phi}$ $(n \times n)$ and $\underset{\sim}{\Psi}(m \times m)$ be the covariance matrices of $\underset{\sim}{\xi}$ and $\underset{\sim}{\zeta}$, respectively, and let $\underset{\sim}{\Theta}_\varepsilon$ and $\underset{\sim}{\theta}_\delta$ be the covariance matrices of $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$, respectively. Then it follows, from the above assumption, that the covariance matrix $\underset{\sim}{\Sigma}[(p + q) \times (p + q)]$ of $\underset{\sim}{z} = (\underset{\sim}{y}', \underset{\sim}{x}')'$ is

$$\Sigma = \begin{bmatrix} \Lambda_y(B^{-1}\Gamma\Phi\Gamma'B'^{-1} + B^{-1}\Psi B'^{-1})\Lambda_y' \quad + & \Lambda_y B^{-1}\Gamma\Phi\Lambda_x' \\ & \\ \Lambda_x\Phi\Gamma'B'^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\varepsilon \end{bmatrix} \qquad (7)$$

The elements of Σ are functions of the elements of Λ_y, Λ_x, B, Γ, Φ, Θ_δ and Θ_ε. In applications some of these elements are fixed and equal to assigned values. In particular, this is so for elements in Λ_y, Λ_x, B and Γ. There is no requirement that $m < p$, $n < q$ and that Θ_ε and Θ_δ be diagonal as in traditional factor analysis. The only requirement is that Σ in (7) is nonsingular and that the model is identified (see section II.E).

There are several options available to the user to choose various special cases of the general model. Probably the most important of these options is the "no x" option, i.e., the specification of a model in which there is no x. Then, the whole equation (6) is missing, so there is no ξ, δ, Λ_x, Φ and Θ_δ. In this case, equation (4) should be interpreted as

$$B\eta = \zeta.$$

Then the only vector of observed variables is y and the covariance matrix of y, $\Sigma(p \times p)$, is

$$\Sigma = \Lambda_y B^{-1}\Psi B'^{-1}\Lambda_y' + \Theta_\varepsilon$$

The only parameter matrices are Λ_y, B, Ψ and Θ_ε.

The measurement model part of the general model, as given by (5) and (6), specifies how the latent variables are measured in terms of the observed variables. This is used to describe the measurement properties (reliabilities and validities) of the observed variables. The structural equation model part of the general model, as given by (4), specifies the causal relationships assumed to hold among the latent variables. This is used to describe and assess the causal effects and to estimate the amount of unexplained variance in the dependent variables. In order to assess the causal effects it is necessary that the units of measurement in the latent variables be defined in a natural way. This can often be done by specifying the unit of measurement to be the same as in one of the observed variables. For further information about LISREL and its uses, see Jöreskog (1973a, 1976) and Jöreskog and Sörbom (1976a-c).

C. COFAMM

The COFAMM model assumes that we have measurements from several independent groups of individuals possibly with different mean vectors and covariance matrices. It is assumed that p variables have been measured in a random sample of individuals from each population.

Let $\mathbf{z}_g$ be a vector of order p, representing the measurements obtained in group g, $g=1,2,\ldots, G$. We regard $\mathbf{z}_g$ as a random vector with mean vector $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$. It is assumed that a factor analysis model holds in each population so that $\mathbf{z}_g$ can be accounted for by k common factors $\mathbf{f}_g$ and p unique factors or residuals $\mathbf{e}_g$, as

$$\mathbf{z}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \mathbf{f}_g + \mathbf{e}_g, \qquad (9)$$

where $\boldsymbol{\nu}_g$ is a pxl vector of location parameters and $\boldsymbol{\Lambda}_g$ a pxk parameter matrix of factor loadings. It is assumed that $E(\boldsymbol{e}_g) = \boldsymbol{0}$ and $E(\boldsymbol{f}_g) = \boldsymbol{\theta}_g$, a kxl parameter vector and that $\boldsymbol{e}_g$ and $\boldsymbol{f}_g$ are uncorrelated. These assumptions imply that the mean vector $\boldsymbol{\mu}_g$ of $\boldsymbol{z}_g$ is

$$\boldsymbol{\mu}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g\boldsymbol{\theta}_g \tag{10}$$

and that the covariance matrix $\boldsymbol{\Sigma}_g$ of $\boldsymbol{z}_g$ is

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g\boldsymbol{\Phi}_g\boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g \tag{11}$$

where $\boldsymbol{\Phi}_g$ is the covariance matrix of $\boldsymbol{f}_g$ and $\boldsymbol{\Psi}_g$ is the covariance matrix of $\boldsymbol{e}_g$.

Concerning the specification of parameters $\boldsymbol{\nu}_g$ and $\boldsymbol{\Lambda}_g$ there are several options. The most important of these is the specification of invariance over groups, i.e.

$$\boldsymbol{\nu}_1 = \boldsymbol{\nu}_2 = \dots = \boldsymbol{\nu}_G$$

$$\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \dots = \boldsymbol{\Lambda}_G.$$

This makes it possible to estimate the $\boldsymbol{\theta}_g$, g=1,2,..., G on a common scale. For further information about COFAMM and its uses, see Jöreskog (1971), Sörbom (1974) and Sörbom and Jöreskog (1976).

D. Fixed, free and constrained parameters

In all three models and computer programs, some elements of any parameter matrix may be fixed and equal to assigned values. For the remaining nonfixed elements of the parameter matrices one or more subsets may have identical but unknown values. Thus each element in any parameter matrix may be:

- (i) *a fixed parameter* that has been assigned a given value,
- (ii) *a constrained parameter* that is unknown but equal to one or more other parameters or
- (iii) *a free parameter* that is unknown and not constrained to be equal to any other parameter.

This results in great generality and flexibility in that many different kinds of models may be handled. The three models and the programs cover a wide range of applications in the behavioral and social sciences.

E. *Identification of parameters*

The general models described here and those that will be considered in later sections of this chapter are all of the following form. The distribution of the observed variables is multivariate with mean vector $\underset{\sim}{\mu}(\underset{\sim}{\theta})$ and covariance matrix $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$ both being functions of parameters $\underset{\sim}{\theta}' = (\theta_1, \theta_2, \ldots, \theta_s)$ which are to be estimated from data. It is assumed that the distribution of the observed variables is sufficiently well described by the moments of first and second order, i.e. by the mean vector $\underset{\sim}{\mu}$ and the covariance matrix $\underset{\sim}{\Sigma}$, so that information about $\underset{\sim}{\theta}$ contained in moments of higher order than the second may be ignored. In particular, this will hold if the distribution is multivariate normal.

In general the parameters in $\underset{\sim}{\theta}$ may be of three kinds

- (i) those that are involved in both $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$
- (ii) those that are involved in $\underset{\sim}{\mu}$ only,
- (iii) those that are involved in $\underset{\sim}{\Sigma}$ only.

Let $\underset{\sim}{\theta}_1$, $\underset{\sim}{\theta}_2$ and $\underset{\sim}{\theta}_3$ be vectors with these three types of parameters, so that $\underset{\sim}{\theta}' = (\underset{\sim}{\theta}_1', \underset{\sim}{\theta}_2', \underset{\sim}{\theta}_3')$. A special case is when $\underset{\sim}{\theta}_1$ is empty as in ACOVSM and LISREL. A further special case is when $\underset{\sim}{\theta}_1$ is empty and the transformation $\underset{\sim}{\mu}(\underset{\sim}{\theta}_2)$ is one-to-one as in LISREL. Then the mean vector $\underset{\sim}{\mu}$ is unconstrained and the only restriction is on $\underset{\sim}{\Sigma}$. Another special case is when $\underset{\sim}{\theta}_1$ is empty and the transformation $\underset{\sim}{\Sigma}(\underset{\sim}{\theta}_3)$ is one-to-one. Then $\underset{\sim}{\Sigma}$ is unconstrained and the only restriction is on $\underset{\sim}{\mu}$.

Before an attempt is made to estimate the parameters $\underset{\sim}{\theta}$, the identification problem must be resolved. The model is said to be identified if $\underset{\sim}{\theta}_1 \neq \underset{\sim}{\theta}_2$ implies that $\left\{\underset{\sim}{\mu}(\underset{\sim}{\theta}_1),\ \underset{\sim}{\Sigma}(\underset{\sim}{\theta}_1)\right\} = \left\{\underset{\sim}{\mu}(\underset{\sim}{\theta}_2),\ \underset{\sim}{\Sigma}(\underset{\sim}{\theta}_2)\right\}$, i.e., if $(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ is generated by one and only one $\underset{\sim}{\theta}$. However, even if the whole model is not identified some parameters in $\underset{\sim}{\theta}$ can still be identified. Consider the set Θ of all $\underset{\sim}{\theta}$ generating the same $(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$. If a parameter θ_i has the same value in all vectors $\underset{\sim}{\theta} \varepsilon \Theta$ then θ_i is said to be identified. For parameters which are identified it is usually possible to find consistent estimators. If a model is not completely identified, restrictions must be imposed on $\underset{\sim}{\theta}$ to make it so. If a parameter is not identified it is not possible to find a consistent estimator of it.

Identifiability depends on the choice of model and on the specification of fixed, constrained and free parameters. To examine the identification problem consider the model equations in the form

$$\left.\begin{aligned} \mu_i &= f_i(\underset{\sim}{\theta}) \\ \sigma_{jk} &= g_{jk}(\underset{\sim}{\theta}),\quad j \leq k \end{aligned}\right\}, \qquad (12)$$

where f_i and g_{jk} are continuous non-linear functions of $\underset{\sim}{\theta}$.

If, for given $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$, a parameter θ can be determined from $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$, this parameter is identified; otherwise it is not. Often some parameters can be determined from $\underset{\sim}{\Sigma}$ and/or $\underset{\sim}{\mu}$ in different ways. This gives rise to overidentifying conditions on $\underset{\sim}{\Sigma}$ and/or $\underset{\sim}{\mu}$ which must hold if the model is true. The solution of (12) is often complicated and tedious and explicit solutions for all θ's seldom exist. It is sometimes difficult to determine whether or not a parameter is identified and whether or not the whole model is identified. Fortunately, however, there is one way in which the computer programs checks the identification status of the model. At the starting point of the iterations, the program computes the information matrix (see e.g., Silvey, 1970) for all the independent unknown parameters. If this matrix is positive definite the model is identified. On the other hand, if the information matrix is singular, the model is not identified. If the information matrix is inverted by the square root method and the n:th pivotal element is zero or negative, this is an indication that the n:th parameter is not identified.

F. Estimation and testing of the models

Once the model has been specified to be of the form suitable for any of the three programs ACOVSM, LISREL and COFAMM, these programs may be used to estimate the model from data. This is done by fitting $\underset{\sim}{\mu}(\underset{\sim}{\theta})$ and $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$ to the corresponding sample estimates $\underset{\sim}{\bar{z}}$, the sample mean vector, and $\underset{\sim}{S}$, the sample covariance matrix. The fitting function is

$$F = N[\log|\Sigma| + \mathrm{tr}(S\Sigma^{-1}) + (\bar{z} - \mu)'\Sigma^{-1}(\bar{z} - \mu) - \log|S| - p]$$

where p is the number of observed variables. F is minimized with respect to θ. This gives maximum likelihood estimates if the distribution of the observed variables is multivariate normal. Standard errors may be obtained for each estimated parameter by computing the inverse of the information matrix at the minimum of F.

The minimum value of F provides a χ^2-goodness-of-fit measure of how well the model fits the data. This may be regarded as a large sample χ^2 test of the specified model against the most general alternative model that both μ and Σ are unconstrained. The degrees of freedom for this test is $(1/2)(p + 1)(p + 2) - s$, where p is the number of observed variables and s is the total number of independent parameters estimated under the model. If μ is unconstrained, the degrees of freedom is $(1/2)p(p + 1) - s$.

Suppose H_o represents one model under given specifications of fixed, free, and constrained parameters. Then it is possible, in large samples, to test the model H_o against any more general model H_1, by estimating each of them separately and comparing their χ^2 goodness-of-fit values. The difference in χ^2 is asymptotically a χ^2 with degrees of freedom equal to the corresponding difference in degrees of freedom. In many situations, it is possible to set up a sequence of hypotheses such that each one is a special case of the preceding and to test these hypotheses sequentially.

In a more exploratory situation the χ^2-goodness-of-fit-values can be used as follows. If a value of χ^2 is obtained, which is large compared to the number of degrees of freedom, the fit may be examined by an inspection of the magnitudes of the first derivatives of F with respect to the fixed parameters. Often such an inspection of the results of analysis will suggest ways to relax the model somewhat by introducing more parameters. The new model usually yields a smaller χ^2. A drop in χ^2 which is large compared to the difference in degrees of freedom, indicates that the changes made in the model represent a real improvement. On the other hand, a drop in χ^2 close to the difference in number of degrees of freedom indicates that the improvement in fit is obtained by "capitalizing on chance" , and the added parameters may not have real significance and meaning.

III. Estimation of Growth Curves

A. One variable over time

Consider one variable y being measured on N individuals at T points in time $t_1, t_2, \ldots, t_T$. The raw data takes the form of a data matrix $\underset{\sim}{Y}$ of order N x T:

$$\begin{bmatrix} y_{11}, y_{12}, \ldots, y_{1T} \\ y_{21}, y_{22}, \ldots, y_{2T} \\ \vdots \\ y_{N1}, y_{N2}, \ldots, y_{NT} \end{bmatrix},$$

where y_{ij} is the observed measurement of individual i at time t_j. We assume that the rows of $\underset{\sim}{Y}$ are independently distributed with the same covariance matrix $\underset{\sim}{\Sigma}$. Also the mean vectors of the rows are assumed to be the same, namely

$$\underset{\sim}{\mu}' = (\mu_1, \mu_2, \ldots, \mu_T)$$

However, in this section, the mean values are not regarded as free parameters, but instead we focus attention to the mean μ_t as a function of t. This gives a growth curve describing how the population mean of y changes over time.

We consider polynomial growth curves of the form

$$\mu_t = \xi_o + \xi_1 t + \xi_2 t^2 + \dots \xi_h t^h, \tag{13}$$

although other mathematical forms may also be considered. The degree of the polynomial h is assumed to be less than or equal to T-1. When h < T-1, the mean vector $\underset{\sim}{\mu}$ is constrained and there is not a one-to-one correspondence between $\mu_1, \mu_2, \dots, \mu_T$ and the polynomial coefficients $\xi_o, \xi_1, \dots, \xi_h$. In this section we consider the estimation of these polynomial coefficients.

Let $\xi_o' = (\xi_o, \xi_1, \dots, \xi_h)$ and let

$$\underset{\sim}{P} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_T \\ t_1^2 & t_2^2 & \dots & t_T^2 \\ \vdots & \vdots & & \vdots \\ t_1^h & t_2^h & \dots & t_T^h \end{bmatrix}.$$

Then the statistical model for the data matrix is

$$E(\underset{\sim}{Y}) = \underset{\sim}{j}\,\underset{\sim}{\xi}'\underset{\sim}{P},$$

where $\underset{\sim}{j}$ is a column vector of order N with all elements equal to one.

If the time points are equidistant it is convenient to use the $h + 1$ first orthogonal polynomials of order T as rows of $\underset{\sim}{P}$ (see e.g. Kendall & Stuart, 1961).

When the rows of $\underset{\sim}{Y}$ have a multinormal distribution, the vector of polynomial coefficients $\underset{\sim}{\xi}$ may be estimated by the maximum likelihood method. The maximum likelihood estimate of $\underset{\sim}{\xi}$ is

$$\hat{\underset{\sim}{\xi}} = (\underset{\sim}{P}\underset{\sim}{S}^{-1}\underset{\sim}{P}')^{-1}\, PS^{-1}\underset{\sim}{\bar{y}} , \qquad (14)$$

where $\underset{\sim}{\bar{y}}$ is the sample mean vector and $\underset{\sim}{S}$ is the sample covariance matrix computed from $\underset{\sim}{Y}$.

The above result generalizes easily to the case of several groups of individuals with possibly different mean vectors. Suppose, for example, that there are two groups with n_1 and n_2 individuals in each group. Let the first n_1 rows of $\underset{\sim}{Y}$ be the measurements on individuals in group 1 and let the last n_2 rows be the measurements on individuals in group 2. The growth curves for the two groups may differ so we assume that there are two distinct growth curves to be estimated, i.e.,

$$E(y_{it}^{(g)}) = \xi_o^{(g)} + \xi_1^{(g)}t + \dots\ \xi_h^{(g)}t^h,\ g = 1,2$$

or in compact form

$$E(\underset{\sim}{Y}) = \underset{\sim}{A}\,\underset{\sim}{\Xi}\,\underset{\sim}{P}, \qquad (15)$$

where

$$\underset{\sim}{A}' = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \end{bmatrix}$$

and

$$\underset{\sim}{\Xi} = \begin{bmatrix} \xi_o^{(1)} & \xi_1^{(1)} & \dots & \xi_h^{(1)} \\ & & & \\ \xi_o^{(2)} & \xi_1^{(2)} & \dots & \xi_h^{(2)} \end{bmatrix}.$$

Let $\underset{\sim}{U} = (1/N)\ \underset{\sim}{A}'\underset{\sim}{A}$, $\underset{\sim}{V} = (1/N)\ \underset{\sim}{A}'\underset{\sim}{Y}$ and $\underset{\sim}{W} = (1/N)\ \underset{\sim}{Y}'\underset{\sim}{Y}$. Then

$$\underset{\sim}{S} = \underset{\sim}{W} - \underset{\sim}{V}'\underset{\sim}{U}^{-1}\underset{\sim}{V} \qquad (16)$$

is the pooled within groups covariance matrix and the maximum likelihood estimate of $\underset{\sim}{\Xi}$ is (see Khatri, 1966)

$$\hat{\underset{\sim}{\Xi}} = \underset{\sim}{U}^{-1}\underset{\sim}{V}\underset{\sim}{S}^{-1}\underset{\sim}{P}'(\underset{\sim}{P}\underset{\sim}{S}^{-1}\underset{\sim}{P}')^{-1} \qquad (17)$$

In general, if there are g independent groups of observations with n_s observations in the s:th group, $n_1 + n_2 + \dots + n_g = N$, the model is still in the form of (15), where $\underset{\sim}{A}$ is of order N x g and has n_1 rows (1, 0, ..., 0), n_2 rows (0, 1, 0, ..., 0), ..., and n_g rows (0, 0, ..., 1). Further,

$$\underset{\sim}{\Xi} = \begin{bmatrix} \xi_o^{(1)} & \xi_1^{(1)} & \dots & \xi_h^{(1)} \\ \xi_o^{(2)} & \xi_1^{(2)} & \dots & \xi_h^{(2)} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \xi_o^{(g)} & \xi_1^{(g)} & & \xi_h^{(g)} \end{bmatrix}$$

and $\underset{\sim}{P}$ is as before. The s:th row of $\underset{\sim}{\Xi}$ consists of the polynomial coefficients for group s. The growth curves are assumed to have the same degree h for all groups. Even in the general case ,the result is given by (17) ,where $\underset{\sim}{U}$, $\underset{\sim}{V}$ and $\underset{\sim}{S}$ are defined as before.

For practical purposes the maximum likelihood estimate of $\underset{\sim}{\Xi}$ may be obtained by means of the computer program ACOVSM (as described in section II). With this program one can also test linear hypotheses on $\underset{\sim}{\Xi}$ of the form

$$\underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{D} = \underset{\sim}{0}$$

where $\underset{\sim}{C}$ (u x g) and $\underset{\sim}{D}$ (h x v) are given matrices of ranks u and v, respectively. In particular, one can test the hypothesis that certain coefficients in one or more growth curves are zero and the hypothesis that certain groups have the same or parallel growth curves. One can also restrict elements of $\underset{\sim}{\Xi}$ to zero in advance. Thus with the ACOVSM program it is not necessary to assume that all groups have growth curves of the same degree. In this case, of course, $\hat{\underset{\sim}{\Xi}}$ is no longer given by (17) but can still easily be computed subject to the zero a priori restrictions.

As an illustration, consider the data in Table 1 taken from Potthoff and Roy (1964). The data is from a dental study in which, the distance, in millimeters, from the center of the pituitary to the pteryomaxillary tissue, was measured on each of 11 girls and 16 boys at ages 8, 10, 12 and 14. The data matrix has 27 rows and 4 columns, the first 11 rows representing the girls and the last 16 rows representing the boys.

Insert Table 1 about here

TABLE 1

Dental Measurements on 11 girls and 16 boys at 4 different ages

Data from Potthoff & Roy (1964), Table 1

Girls					Boys				
	Age in Years					Age in Years			
Individual	8	10	12	14	Individual	8	10	12	14
1	21	20	21.5	23	1	26	25	29	31
2	21	21.5	24	25.5	2	21.5	22.5	23	26.5
3	20.5	24	24.5	26	3	23	22.5	24	27.5
4	23.5	24.5	25	26.5	4	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	5	20	23.5	22.5	26
6	20	21	21	22.5	6	24.5	25.5	27	28.5
7	21.5	22.5	23	25	7	22	22	24.5	26.5
8	23	23	23.5	24	8	24	21.5	24.5	25.5
9	20	21	22	21.5	9	23	20.5	31	26
10	16.5	19	19	19.5	10	27.5	28	31	31.5
11	24.5	25	28	28	11	23	23	23.5	25
					12	21.5	23.5	24	28
					13	17	24.5	26	29.5
					14	22.5	25.5	25.5	26
					15	23	24.5	26	30
					16	22	21.5	23.5	25
Mean	21.18	22.23	23.09	24.09	Mean	22.87	23.81	25.72	27.47

The following two questions may be asked.

(i) Should the growth curves be represented by second degree polynomials or are linear equations adequate?

(ii) Should two separate growth curves be used for boys and girls, or do both have the same growth curve?

To answer these questions we set up a model as in (15) with N=27, T=4, and g=2 and h=2, with $\underset{\sim}{A}$,a matrix of order N x 2 with the first 11 rows equal to (1, 0) and the last 16 rows equal to (0, 1) and with

$$\underset{\sim}{\Xi} = \begin{bmatrix} \xi_0^{(G)} & \xi_1^{(G)} & \xi_2^{(G)} \\ \xi_0^{(B)} & \xi_1^{(B)} & \xi_2^{(B)} \end{bmatrix} .$$

Since the time points are equidistant we take the rows of $\underset{\sim}{P}$ as the first three orthogonal polynomials of order four, i.e.,

$$\underset{\sim}{P} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \\ 9 & 1 & 1 & 9 \end{bmatrix}$$

The maximum likelihood estimate of $\underset{\sim}{\Xi}$ is

$$\hat{\underset{\sim}{\Xi}} = \begin{bmatrix} 22.704 & 0.479 & -0.003 \\ 24.631 & 0.788 & 0.050 \end{bmatrix}$$

To examine question (i) we test the hypothesis $\xi_2^{(G)} = \xi_2^{(B)} = 0$. This can be done by choosing

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

The test statistic can be transformed to an F distribution (see Pottoff & Roy, 1964). In this case one obtains an F=1.19 with 2 and 24 degrees of freedom. This indicates that the hypothesis cannot be rejected. We may therefore regard the growth curves as linear rather than quadratic.

We now modify the model and take $\underset{\sim}{\Xi}$ as

$$\underset{\sim}{\Xi} = \begin{bmatrix} \xi_0^{(G)} & \xi_1^{(G)} \\ \xi_0^{(B)} & \xi_1^{(B)} \end{bmatrix}$$

and $\underset{\sim}{P}$ with only two rows instead of three. The maximum likelihood estimate of $\underset{\sim}{\Xi}$ is now

$$\hat{\underset{\sim}{\Xi}} = \begin{bmatrix} 22.689 & 0.477 \\ 24.923 & 0.826 \end{bmatrix}$$

To examine question (ii) we set up the hypothesis $\xi_0^{(G)} = \xi_0^{(B)}$, $\xi_1^{(G)} = \xi_1^{(B)}$. This corresponds to choosing

$$\underset{\sim}{C} = \begin{bmatrix} 1, & -1 \end{bmatrix}, \qquad \underset{\sim}{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Also, in this case, the test statistic can be transformed to an F distribution. The test gives F = 6.44 with 2 and 22 degrees of freedom, suggesting that the hypothesis should be rejected. Boys and girls have different growth curves.

B. An autoregressive model for one variable over time

The development in the previous subsection did not take the covariance structure in $\underset{\sim}{\Sigma}$ into account. However, the growth curves can be estimated more efficiently and the tests will be more powerful if the covariance structure, which arises naturally in repeated measurements, is taken into account. This covariance structure very often has an autoregressive nature. Therefore, in this subsection, we focus attention to the deviation $e_t = y_t - \mu_t$ of y_t from its mean value μ_t on the growth curve and consider various autoregressive models for this.

The first-order autoregressive model is

$$e_t = \beta_t e_{t-1} + z_t , \quad t = 2, 3, \ldots, T , \qquad (18)$$

where the residual z_t is uncorrelated with e_{t-1}. It is also assumed that $z_2, z_3, \ldots, z_T$ are all uncorrelated. A path diagram of this model is shown in Fig. 2 for the case of T=4.

It is readily verified that

$$\mathrm{Cov}\ (y_t, y_{t-1}) = E(e_t \ \ e_{t-1}) = \beta_t \sigma^2_{t-1},$$

where $\sigma^2_{t-1} = \mathrm{Var}(y_{t-1}) = E(e^2_{t-1})$, and that

$$\mathrm{Cov}\ (y_t, y_{t-k}) = \beta_t \beta_{t-1} \cdots \beta_{t-k+1}\, \sigma^2_{t-k}, \quad k=1,2,\ldots$$

where $\sigma^2_{t-k} = \mathrm{Var}\ (y_{t-k})$.

Insert Figure 2 about here

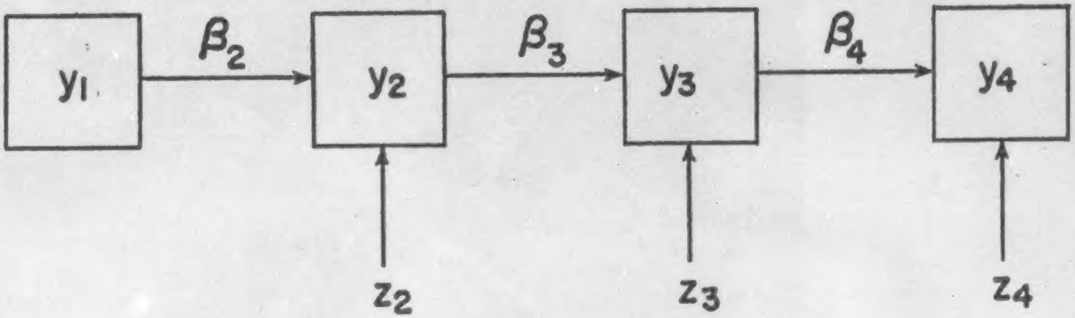

y_1
β_2
y_2
β_3
y_3
β_4
y_4
z_2
z_3
z_4

Figure 2 -- An autoregressive model for one variable over time.

Hence the covariance matrix of $\underset{\sim}{y}$ is (in the case of T = 4)

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_1^2 & & & \\ \beta_2\sigma_1^2 & \sigma_2^2 & & \\ \beta_3\beta_2\sigma_1^2 & \beta_3\sigma_2^2 & \sigma_3^2 & \\ \beta_4\beta_3\beta_2\sigma_1^2 & \beta_4\beta_3\sigma_2^2 & \beta_4\sigma_3^2 & \sigma_4^2 \end{bmatrix} . \qquad (19)$$

From (18) it is seen that $\underset{\sim}{\Sigma}$ is constrained; its ten variances and covariances are functions of only seven parameters. Since the variances are free parameters it is the six covariances that are functions of the three parameters β_2, β_3 and β_4. In the general case there are (1/2)T (T + 1) variances and covariances in $\underset{\sim}{\Sigma}$ and 2T - 1 free parameters.

The correlation matrix corresponding to (19) is

$$\begin{bmatrix} 1 & & & \\ \rho_2 & 1 & & \\ \rho_2\rho_3 & \rho_3 & 1 & \\ \rho_2\rho_3\rho_4 & \rho_3\rho_4 & \rho_4 & 1 \end{bmatrix} , \qquad (20)$$

where $\rho_i = \beta_i(\sigma_{i-1}/\sigma_i)$. There are only T-1 independent correlations, namely those just below (or above) the main diagonal and the other correlations are products of these. For example,

$$\rho_{ji} = \prod_{k=i+1}^{k=j} \rho_k \qquad (i < j). \tag{21}$$

It is seen that the correlations fall off as one moves away from the main diagonal, a phenomenon usually found to occur empirically. The partial correlation $\rho_{ik \cdot j}$ is zero, whenever $i < j < k$. This is readily verified since $\rho_{ik} - \rho_{ij}\rho_{jk} = 0$ by virtue of (21). Higher order partial correlations, with two or more intermediate variables held constant, also vanish. It follows that in the regression

$$y_t = \beta_{t1}y_1 + \beta_{t2}y_2 + \dots + \beta_{t,t-1}y_{t-1} + \ddot{z}_t \tag{22}$$

of y_t on all preceding variables, the only regression coefficient that can be non-zero is $\beta_{t,t-1} = \beta_t$. For the prediction of y_t, only the immediate neighbor y_{t-1} is useful. The effects of y_{t-2}, y_{t-3}, ..., y_1 on y_t are only indirect via y_{t-1}.

If the growth curve specification in (15) is ignored so that the mean vector $\underset{\sim}{\mu}$ is unconstrained, this model can be estimated very easily. Under multinormality, the maximum likelihood estimate of β_t is just the ordinary least squares estimate one obtains by estimating each regression equation in (18) separately, namely

$$\hat{\beta}_t = s_{t-1,t}/s_{t-1,t-1} ,$$

where the s_{ij} are elements of $\underset{\sim}{S}$ in (16). The residual variance, $\mathrm{Var}(z_t)$, is estimated as

$$\mathrm{Var}(z_t) = s_{tt} - \hat{\beta}_t^2 s_{t-1,t-1} .$$

The estimation of the growth curves and the β's simultaneously is not so easy. This may be done numerically by means of the ACOVSM program. We now show that the $\underset{\sim}{\Sigma}$ in (19) is indeed of the form required by that program. To do so we define $z_1 \equiv e_1$ and write the equations (18) as (in the case of T = 4)

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta_2 & 1 & 0 & 0 \\ \beta_2\beta_3 & \beta_3 & 1 & 0 \\ \beta_2\beta_3\beta_4 & \beta_3\beta_4 & \beta_4 & 1 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} . \qquad (23)$$

Let $\kappa_i = \beta_2\beta_3 \ldots \beta_i$ for i = 2, 3, ..., T. Then if all $\beta_i \neq 0$ there is a one-to-one correspondence between $\kappa_{.2}, \kappa_{.3}, \ldots, \kappa_{.T}$ and $\beta_2, \beta_3, \ldots, \beta_T$ and $\beta_i = \kappa_{.i}/\kappa_{i-1}$ The matrix in (23) is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \kappa_2 & 1 & 0 & 0 \\ \kappa_3 & \kappa_3/\kappa_2 & 1 & 0 \\ \kappa_4 & \kappa_4/\kappa_2 & \kappa_4/\kappa_3 & 1 \end{bmatrix} = \underset{\sim}{D}_\kappa \underset{\sim}{T} \underset{\sim}{D}_\kappa^{-1} ,$$

where $\underset{\sim}{D}_\kappa =$ diag $(1, \kappa_2, \kappa_3, \kappa_4)$ and

$$\underset{\sim}{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Hence, (23) can be written

$$\begin{aligned}\underset{\sim}{e} &= \underset{\sim}{D}_{\kappa}\underset{\sim}{T}\underset{\sim}{D}_{\kappa}^{-1}\underset{\sim}{z} \\ &= \underset{\sim}{D}_{\kappa}\underset{\sim}{T}\underset{\sim}{z}^{*} ,\end{aligned}$$

with covariance matrix

$$\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\kappa}\underset{\sim}{T}\underset{\sim}{\Psi}^{*}\underset{\sim}{T}'\underset{\sim}{D}_{\kappa} , \qquad (24)$$

where $\underset{\sim}{\Psi}^{*}$ is the diagonal covariance matrix of $\underset{\sim}{z}^{*} = \underset{\sim}{D}_{\kappa}^{-1}\underset{\sim}{z}$. The $2T - 1$ parameters $\kappa_2, \kappa_3, \ldots, \kappa_T, \psi_{11}^{*}, \psi_{22}^{*}, \ldots, \psi_{TT}^{*}$ are in a one-to-one correspondence with the original parameters $\beta_2, \beta_3, \ldots, \beta_T, \sigma_{11}, \sigma_{22}, \ldots, \sigma_{TT}$. Equation (24) is in the form of (2). The whole model is defined by (15) and (24). The ACOVSM program gives estimates of the growth curve polynomial coefficients as well as $\beta_2, \beta_3, \ldots, \beta_T, \sigma_{11}, \sigma_{22}, \ldots, \sigma_{TT}$. The program also gives a χ^2-goodness-of-fit-measure for assessing the fit of the overall model. This χ^2-measure may be divided into two components measuring the fit of the growth curve model (15) and the covariance structure model (24) separately.

C. Growth Curves for several variables simultaneously

The model of the preceding subsection will now be generalized to the case of several variables at each occasion. We still assume that the observed variables are measured without errors. The case of measurement errors in the dependent variables will be considered in sections IV and V. An example of the type of model to be considered is shown in Fig. 3. Here there are 3 variables for all t and, as before, we illustrate with T = 4 occasions.

Insert Figure 3 about here

The growth curve specification for the model in Fig. 3 is as follows. For an arbitrary individual we arrange his observed scores so that his three scores at the first occasion come first, then his three scores at the second occasion, etc., i.e.,

$$y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}, y_{31}, y_{32}, y_{33}, y_{41}, y_{42}, y_{43},$$

where y_{tj} is the score on variable j at occasion t. The growth curve for variable j is assumed to be

$$E(y_{tj}) = \xi_{jo} + \xi_{j1}t + \xi_{j2}t^2 , \qquad (25)$$

say. As before, the model is given in matrix form by (15), where

$$\underset{\sim}{\Xi}\ (1 \times 6) = (\xi_{10}, \xi_{11}, \xi_{12}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{30}, \xi_{31}, \xi_{32})$$

and

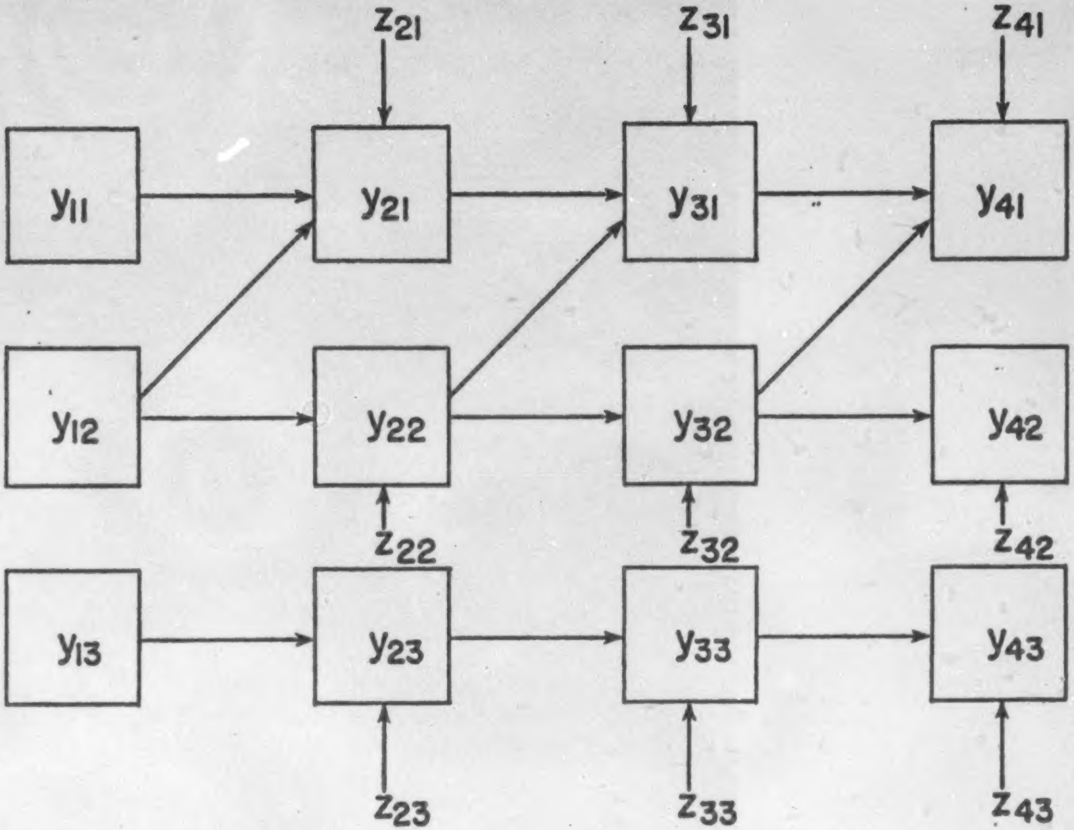

z21
z31
z41
y11
y21
y31
y41
y12
y22
y32
y42
z22
z32
z42
y13
y23
y33
y43
z23
z33
z43

Figure 3 -- A multivariate autoregressive model.

$$P(6 \times 8) = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ t_1 & 0 & t_2 & 0 & t_3 & 0 & t_4 & 0 \\ t_1^2 & 0 & t_2^2 & 0 & t_3^2 & 0 & t_4^2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & t_1 & 0 & t_2 & 0 & t_3 & 0 & t_4 \\ 0 & t_1^2 & 0 & t_2^2 & 0 & t_3^2 & 0 & t_4^2 \end{bmatrix}$$

The matrix A is a column vector of order N with all elements equal to one. If there are g groups of observations, there will be g rows in Ξ and the matrix A will as be described in section III.A.

Without constraints on the covariance matrix Σ and with no a priori zero restrictions on Ξ, this model can be estimated as in III.A. The maximum likelihood estimate of Ξ is given by (17).

D. Multivariate autoregressive models

We now consider a multivariate autoregressive model which is a direct generalization of the univariate autoregressive model in section III.B. This autoregressive model is

$$y_t = B_t y_{t-1} + z_t , \quad t = 2, 3, \ldots, T, \tag{26}$$

with the y's measured as deviations from their means. For the model in Fig. 3, each matrix B_t will be of the form

$$\mathbf{B}_t = \begin{bmatrix} \beta_{11}^{(t)} & \beta_{12}^{(t)} & 0 \\ 0 & \beta_{22}^{(t)} & 0 \\ 0 & 0 & \beta_{33}^{(t)} \end{bmatrix} \qquad (27)$$

The residuals in $\mathbf{z}_t$ are assumed to be uncorrelated across time but may be contemporaneously correlated, i.e., $E(\mathbf{z}_t\mathbf{z}_s') = \mathbf{0}$ for $s \neq t$. The covariance matrix $E(\mathbf{z}_t\mathbf{z}_t')$ is denoted $\mathbf{\Phi}_t$. If the mean vectors $\boldsymbol{\mu}_t$ and the matrices $\mathbf{B}_t$ are unconstrained, this model may be estimated directly by estimating each regression in (26) separately. Let

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1T} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2T} \\ \vdots & & & \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} & \cdots & \mathbf{S}_{TT} \end{bmatrix}$$

be the sample covariance matrix of $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_t')$, where $\mathbf{S}_{st}$ is the covariance matrix between $\mathbf{y}_t$ and $\mathbf{y}_s$. Then the maximum likelihood estimates are

$$\hat{\mathbf{B}}_t = \mathbf{S}_{t,t-1}\mathbf{S}^{-1}_{t-1,t-1}$$

and

$$\hat{\mathbf{\Psi}}_t = \mathbf{S}_{tt} - \mathbf{S}_{t,t-1}\,\mathbf{S}^{-1}_{t-1,\,t-1}\,\mathbf{S}_{t-1,t}\ .$$

If the matrices $\underset{\sim}{B}_t$ have fixed zero elements as in (27), the model may be estimated by means of the LISREL program as described in section 2. LISREL can estimate the covariance structure but not the growth curves specification for the means. When the latter are included in the model together with the multivariate autoregressive model, the estimation problem is complicated and there does not seem to be any general program available to handle this estimation. When all $\underset{\sim}{B}_t$ are diagonal, ACOVSM may be used in the same way as in III.B.

IV. Two-Wave Models

A. Two-wave, two-variable models

In the previous section all the variables were assumed to be measured without error. Measurement errors in the variables were not taken into account in the models which focused directly on relationships between the observed variables. In this and the next section we assume that all the observed variables contain errors of measurement and focus on the relationships among the true or latent variables. In doing so we shall ignore any structure on the mean vector and simply assume this to be unconstrained. We may therefore take all variables to be measured in deviations from their means.

We begin with the simple model shown in Fig. 4, where two variables are measured at two occasions. We assume that the two variables measures the same latent variable η, i.e., y_{11} and y_{12} measures η_1 on the first occasion and y_{21} and y_{22} measures η_2 on the second occasion. We are interested in the relationship between η_1 and η_2 expressed in the structural equation

$$\eta_2 = \beta\eta_1 + \zeta, \tag{28}$$

the regression of η_2 on η_1. In particular, we are interested in whether $\beta = 1$ and ζ is small, i.e., whether the same latent variables are measured on both occasions.

Insert Figure 4 about here

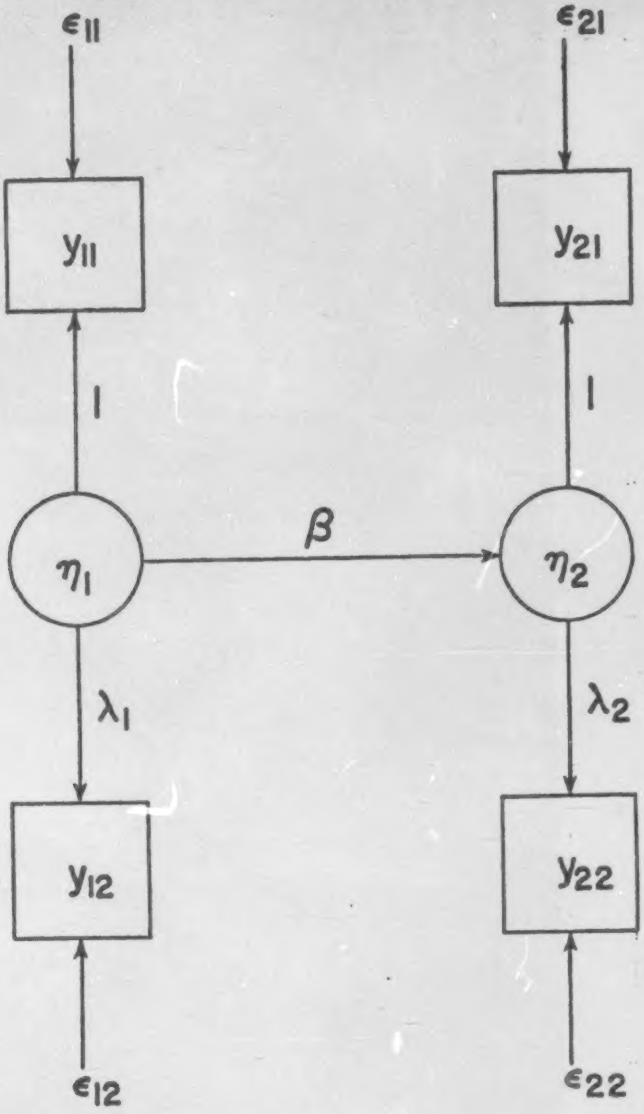

ϵ_{11}
ϵ_{21}
y_{11}
y_{21}
1
1
η_1
β
η_2
λ_1
λ_2
y_{12}
y_{22}
ϵ_{12}
ϵ_{22}

Figure 4 -- A two-wave, two-variable model.

The measurement model part of the model may be written as

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix} , \qquad (29)$$

where it is assumed that η_1 and η_2 are measured in the same metric as y_{11} and y_{21}, respectively. This model is a special case of the general LISREL model with no $\underset{\sim}{x}$. In terms of LISREL, (28) may be interpreted, in accordance with (4), as

$$\begin{pmatrix} 1 & 0 \\ -\beta & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} ,$$

where $\zeta_1 = \eta_1$ and $\zeta_2 = \zeta$. Let $\underset{\sim}{\Phi}$ be the covariance matrix of (η_1, η_2) and let $\underset{\sim}{\Theta}$ be the covariance matrix of $(\varepsilon_{11}, \varepsilon_{12}, \varepsilon_{21}, \varepsilon_{22})$. If all the ε's are uncorrelated so that $\underset{\sim}{\Theta}$ is diagonal, the covariance matrix of $(y_{11}, y_{12}, y_{21}, y_{22})$ is

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \phi_{11} + \theta_{11} & & & \\ \lambda_1\phi_{11} & \lambda_1^2\phi_{11} + \theta_{22} & & \\ \phi_{21} & \lambda_1\phi_{21} & \phi_{22} + \theta_{33} & \\ \lambda_2\phi_{21} & \lambda_1\lambda_2\phi_{21} & \lambda_2\phi_{22} & \lambda_2^2\phi_{22} + \theta_{44} \end{bmatrix} ,$$

$\underset{\sim}{\Sigma}$ has 10 variances and covariances which are functions of 9 parameters. The model has one degree of freedom.

Often when the same variables are used repeatedly there is a tendency for the corresponding errors (the ε's) to correlate over time (see section III.B and III.D) because of memory and other retest effects. Hence there is a need to generalize the preceding model to allow for correlations between ε_{11} and ε_{21} and also between ε_{12} and ε_{22}. This means that there will be two non-zero covariances θ_{31} and θ_{42} in $\underset{\sim}{\Theta}$. This model is shown in Fig. 5. The covariance matrix of the observed variables will now be

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \phi_{11} + \theta_{11} & & & \\ \lambda_1\phi_{11} & \lambda_1^2\phi_{11} + \theta_{22} & & \\ \phi_{21} + \theta_{31} & \lambda_1\phi_{21} & \phi_{22} + \theta_{33} & \\ \lambda_2\phi_{21} & \lambda_1\lambda_2\phi_{21} + \theta_{42} & \lambda_2\phi_{22} & \lambda_2^2\phi_{22} + \theta_{44} \end{bmatrix} .$$

This $\underset{\sim}{\Sigma}$ has its 10 independent elements expressed in terms of 11 parameters. Hence it is clear that the model is not identified. In fact, none of the 11 parameters are identified without further restrictions. The loading λ_1 and λ_2 may be multiplied by a constant and the ϕ's divided by the same constant. This does not change σ_{21}, σ_{32}, σ_{41} and σ_{43}. The change in the other σ's may be compensated by adjusting the θ's additively. Hence to make the model identified one must fix one λ or one ϕ at a non-zero value

----

Insert Figure 5 about here

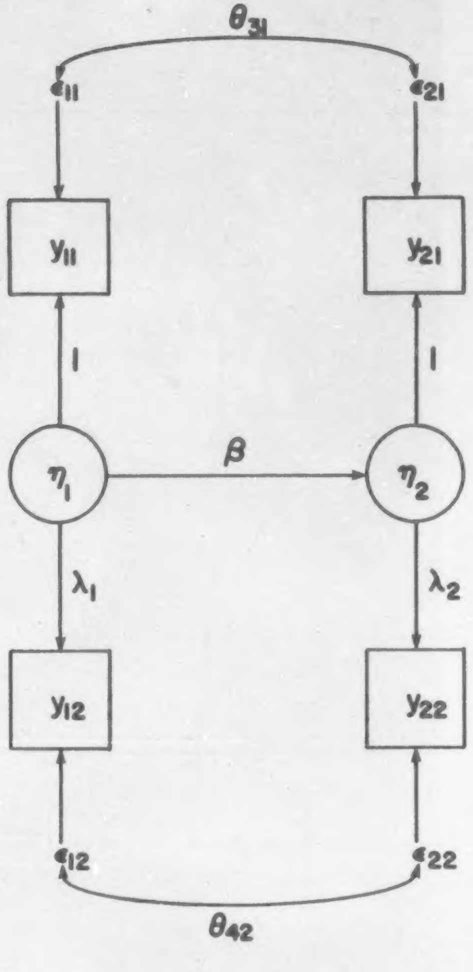

θ_{31}
ϵ_{11}
ϵ_{21}
y_{11}
y_{21}
1
1
η_1
β
η_2
λ_1
λ_2
y_{12}
y_{22}
ϵ_{12}
ϵ_{22}
θ_{42}

Figure 5 -- A two-wave, two-variable model with correlated errors.

or one θ at some arbitrary value. However, the *correlation* between η_1 and η_2 is identified without any restrictions, since

$$\mathrm{Corr}(\eta_1,\eta_2) = \sqrt{\phi_{21}^2/\phi_{11}\phi_{22}}$$

$$= \sqrt{(\sigma_{32}\sigma_{41})/(\sigma_{21}\sigma_{43})}$$

This model may therefore be used to estimate this correlation coefficient and to test whether this is one. The maximum likelihood estimate of the correlation coefficient is $\sqrt{(s_{32}s_{41})/(s_{21}s_{43})}$. To make further use of the model it is necessary to make some assumption about the nature of the variables. For example, if it can be assumed that the two variables at each occasion are tau-equivalent (see e.g. Lord & Novick, 1968) we can set both λ_1 and λ_2 equal to one. Then the model can be estimated and tested with one degree of freedom.

B. Two-wave, two-variables models with background variables

The model of the previous subsection may be used for the measurement of change between two occasions. However, in many longitudinal studies the objective is not only to measure change but also to attribute or relate change to certain characteristics and events. Such studies must include not only pre- and postmeasures but also various background variables believed to influence change. The background variables may be socioeconomic variables or other characteristics differentiating the individuals prior to the pretest occasion.

Consider the model shown in Fig. 6. The background variable is denoted x. The main purpose of the model is to separate the direct effect of η_1 on η_2 by eliminating the effect of x.

The measurement model for $\underset{\sim}{y}$ is the same as in (29) but now the structural equations are

$$\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} x + \begin{pmatrix} \zeta_2 \\ \zeta_2 \end{pmatrix} . \qquad (30)$$

The $\underset{\sim}{\Lambda}_x$ in (6) is a 1 x 1 matrix with element one and $\underset{\sim}{\delta} = 0$.

The reduced form of (30) is

$$\begin{aligned} \eta_1 &= \gamma_1 x + \zeta_1 \\ \eta_2 &= (\gamma_2 - \beta\gamma_1)x + (\zeta_2 - \beta\zeta_1) \\ &= \pi \, x + \upsilon \text{ , say.} \end{aligned}$$

Insert Figure 6 about here

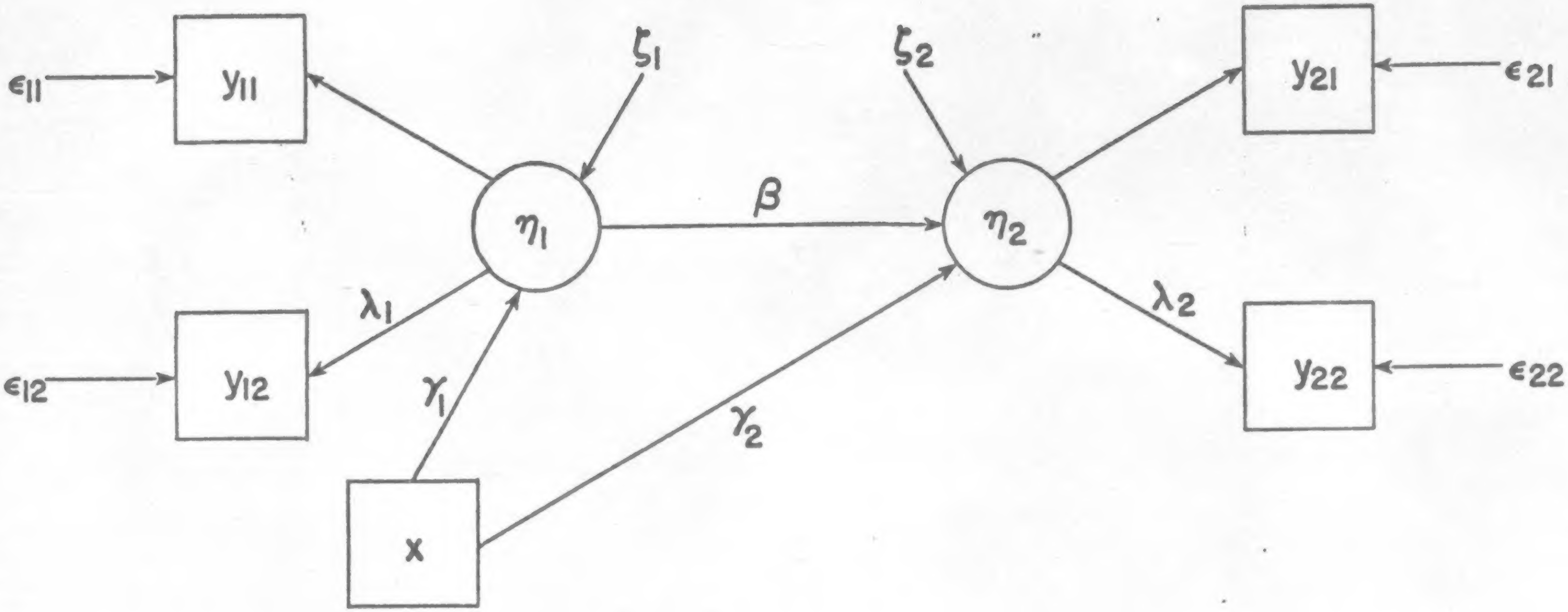

ϵ_{11}
y_{11}
ϵ_{12}
y_{12}
λ_1
ζ_1
η_1
γ_1
x
β
γ_2
ζ_2
η_2
λ_2
y_{21}
ϵ_{21}
y_{22}
ϵ_{22}

Figure 6 -- A two-wave, two-variable model with an infallible background variable.

As before, we assume that the measurement errors ε_{11}, ε_{12}, ε_{21} and ε_{22} have zero means and are mutually uncorrelated and uncorrelated with η_1 and η_2. Furthermore, we assume that the residuals ζ_1 and ζ_2 have zero means and are uncorrelated. The variances of ζ_1 and ζ_2 are denoted $\psi_{11} = \mathrm{Var}(\zeta_1)$, $\psi_{22} = \mathrm{Var}(\zeta_2)$.

Let us first consider the identification problem. We have five observed variables y_{11}, y_{12}, y_{21}, y_{22} and x with fifteen variances and covariances. The model has the following twelve parameters to be estimated λ_1, λ_2, β, γ_1, γ_2, $\phi = \mathrm{Var}(x)$, ψ_{11}, ψ_{22} and θ_{ii}, $i = 1, 2, 3, 4$. We have

$$\mathrm{Cov}(y_{11},x) = \mathrm{Cov}(\eta_1,x) = \gamma_1\phi$$

$$\mathrm{Cov}(y_{12},x) = \lambda_1\mathrm{Cov}(\eta_1,x) = \lambda_1\gamma_1\phi$$

$$\mathrm{Cov}(y_{21},x) = \mathrm{Cov}(\eta_2,x) = \pi\phi$$

$$\mathrm{Cov}(y_{22},x) = \lambda_2\mathrm{Cov}(\eta_2,x) = \lambda_2\pi\phi \quad .$$

Since $\phi = \mathrm{Var}(x)$ is identified, these equations determine γ_1, λ_1, π and λ_2, respectively. Furthermore,

$$\mathrm{Cov}(y_{11},y_{12}) = \lambda_1\mathrm{Var}(\eta_1) = \lambda_1(\gamma_1^2\phi + \psi_{11}),$$

which determines ψ_{11} , and

$$\mathrm{Cov}(y_{21}, y_{22}) = \lambda_2\mathrm{Var}(\eta_2) = \lambda_2[\pi^2\phi + \mathrm{Var}(\upsilon)],$$

which determines

$$\mathrm{Var}(\upsilon) = \psi_{22} + \beta^2\psi_{11}. \tag{31}$$

For given λ_1, λ_2, γ_1, π, ϕ and ψ_{11}, any one of the four equations

$$\mathrm{Cov}(y_{11},y_{21}) = \gamma_1\pi\phi - \beta\psi_{11} \ , \tag{32}$$

$$\mathrm{Cov}(y_{11},y_{22}) = \lambda_2(\gamma_1\pi\phi - \beta\psi_{11}) \ , \tag{33}$$

$$\mathrm{Cov}(y_{12},y_{21}) = \lambda_1(\gamma_1\pi\phi - \beta\psi_{11}), \tag{34}$$

$$\mathrm{Cov}(y_{12},y_{22}) = \lambda_1\lambda_2(\gamma_1\pi\phi - \beta\psi_{11}) \ , \tag{35}$$

determine β . Then, with β determined, $\gamma_2 = \pi + \beta\gamma_1$ and ψ_{22} is obtained from (31). The error variances θ_{ii} are determined from σ_{ii}, $i = 1,2,3,4$. Hence it is clear that the whole model is identified and has three independent restrictions on Σ.

Now suppose that x cannot be measured without error and write

$$x = \xi + \delta \ ,$$

where ξ is the true score and δ the measurement error, the latter assumed to have zero mean and to be uncorrelated with ξ and everything else. We shall consider two cases namely: (a) x has a known reliability $\rho_{xx} = \sigma^2_\xi / \sigma^2_x$ and (b) ξ is measured by two congeneric background variables x_1 and x_2. Case (a) is shown in Fig. 7.

In case (a), the above equations are the same except that ϕ is replaced by σ^2_ξ . Since $\sigma^2_\xi = \rho_{xx}\phi$ where ρ_{xx} is known and ϕ is identified, all the other parameters will be determined as before.

Insert Figure 7 about here

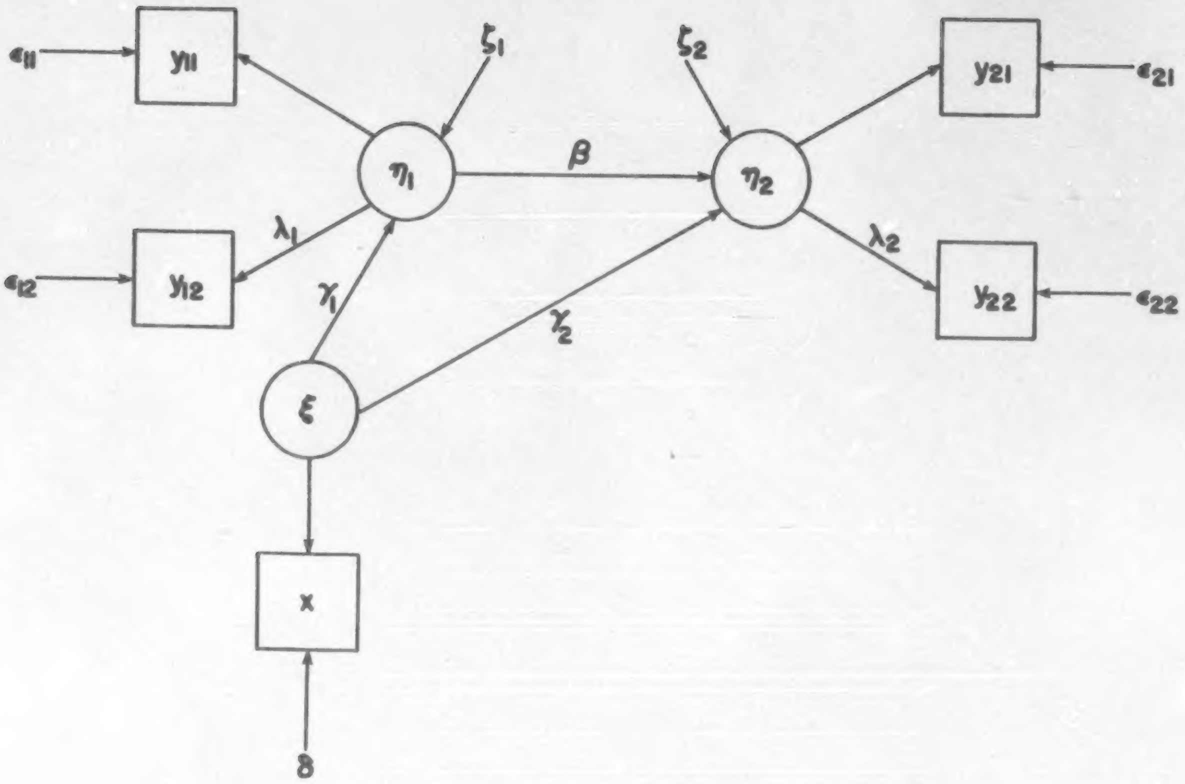

ϵ_{11}
y_{11}
ζ_1
ζ_2
y_{21}
ϵ_{21}
η_1
β
η_2
λ_1
λ_2
ϵ_{12}
y_{12}
γ_1
γ_2
y_{22}
ϵ_{22}
ξ
x
δ

Figure 7 -- A two-wave, two-variable model with a fallible background variable.

Next suppose the errors ε_{11} and ε_{21} and also ε_{12} and ε_{22}, are correlated as in section IV.A. Such a model is shown in Fig. 8. Then θ_{31} will be added to the right side of (32) and θ_{42} will be added to the right side of (35). Equations (33) and (34) still determine β for given λ_1, λ_2, γ_1, π, ϕ and ψ_{11} and θ_{31} and θ_{42} are then determined by (32) and (35), respectively. Hence this model has one overidentifying restriction.

Case (b) is shown in Fig. 9. Here we write

$$x_1 = \xi + \delta_1$$

$$x_2 = \lambda_3\xi + \delta_2$$

where λ_3 is a parameter to be determined and δ_1 and δ_2 are uncorrelated measurement errors, uncorrelated with ξ and the other latent variables. The other equations are as before except that x is replaced by ξ. We then have three more parameters than before namely λ_3, $\sigma^2_{\delta 1}$ and $\sigma^2_{\delta 2}$. The parameter $\sigma^2_\xi = \mathrm{Var}(\xi)$ replaces $\phi = \sigma^2_x = \mathrm{Var}(x)$. On the other hand we have now six more manifest parameters, so that the model has six degrees of freedom with $\theta_{31} = \theta_{42} = 0$ and four degrees of freedom with these covariances included as parameters.

The parameter λ_3 is identified with three overidentifying restrictions since

$$\mathrm{Cov}(x_2,w)/\mathrm{Cov}(x_1,w) = \lambda_3,$$

for $w = y_{11}$, y_{12}, y_{21} and y_{22}. All the other parameters are determined as before.

Insert Figures 8 and 9 about here

The models in Figs. 6-9 have deliberately been chosen simple to explicate the principal points. The models can easily be generalized in two ways. Firstly, the number of pre- and post measures $\underset{\sim}{y}$ can be more than two. Secondly, we could also have several background variables with a factor structure. We now give two examples of models of this kind.

C. The Stability of Alienation

For the first example we draw on ideas and data in Wheaton et al (1977). Their study was concerned with the stability over time of attitudes such as alienation and its relation to background variables such as education and occupation. Data on attitude scales were collected from 932 persons in two rural regions in Illinois at three points in time: 1966, 1967 and 1971. (See Summers et al, 1969 for further description of the research setting.) The variables we use for the present illustration are the *Anomia* subscale and the *Powerlessness* subscale, taken to be indicators of *Alienation*. We use these subscales from 1967 and 1971 only. The background variables are respondent's education (years of schooling completed) and Duncan's Socioeconomic Index (SEI). These are taken to be indicators of respondent's socioeconomic status (SES). We analyze these variables under three different models as shown in Figures 10A-C none of which correspond to that of Wheaton et al (1977). The data are given in Table 2.

Insert Figure 10 and Table 2 about here

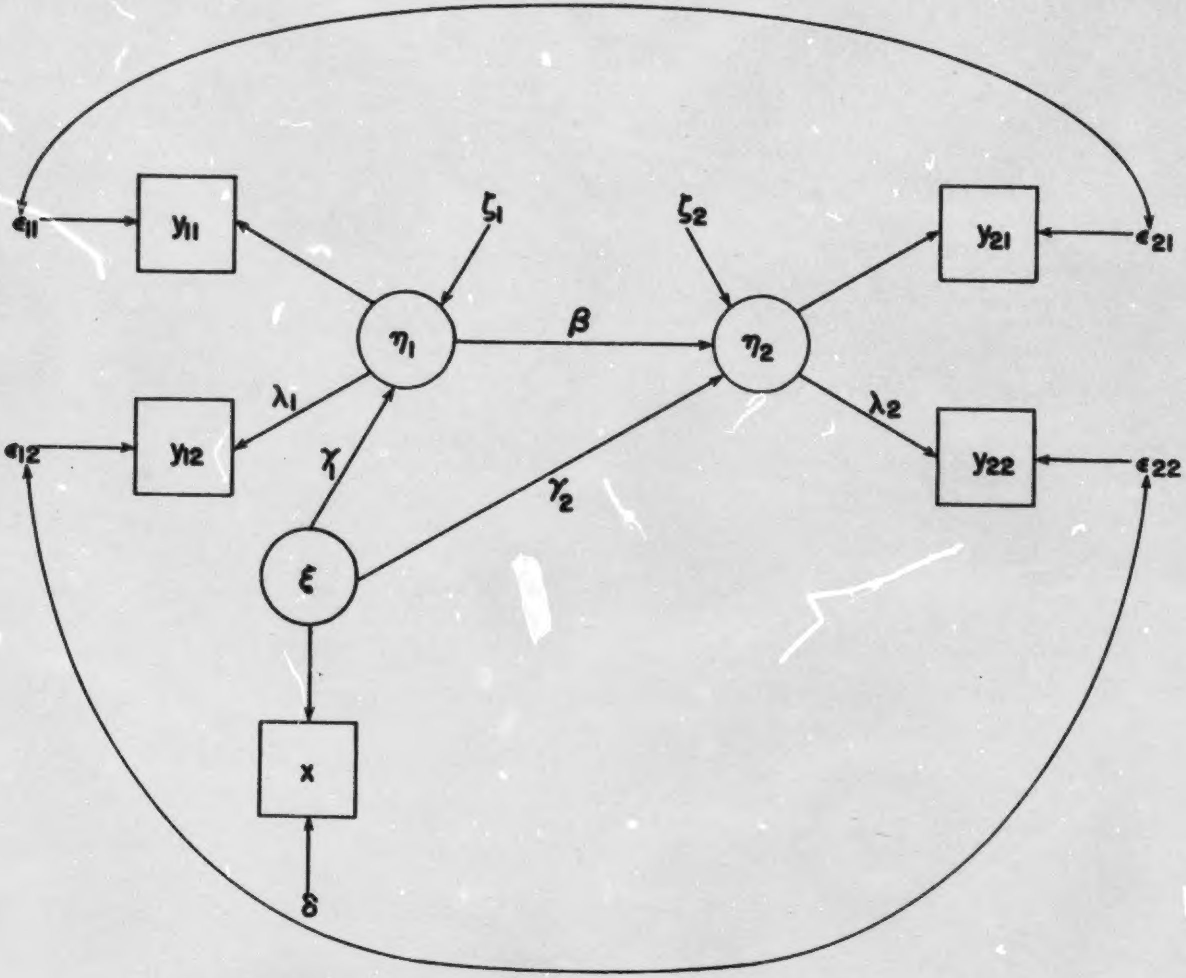

ϵ_{11}
y_{11}
ζ_1
ζ_2
y_{21}
ϵ_{21}
η_1
β
η_2
λ_1
λ_2
ϵ_{12}
y_{12}
y_{22}
ϵ_{22}
γ_1
γ_2
ξ
x
δ

Figure 8 -- A two-wave, two-variable model with correlated errors and a fallible background variable.

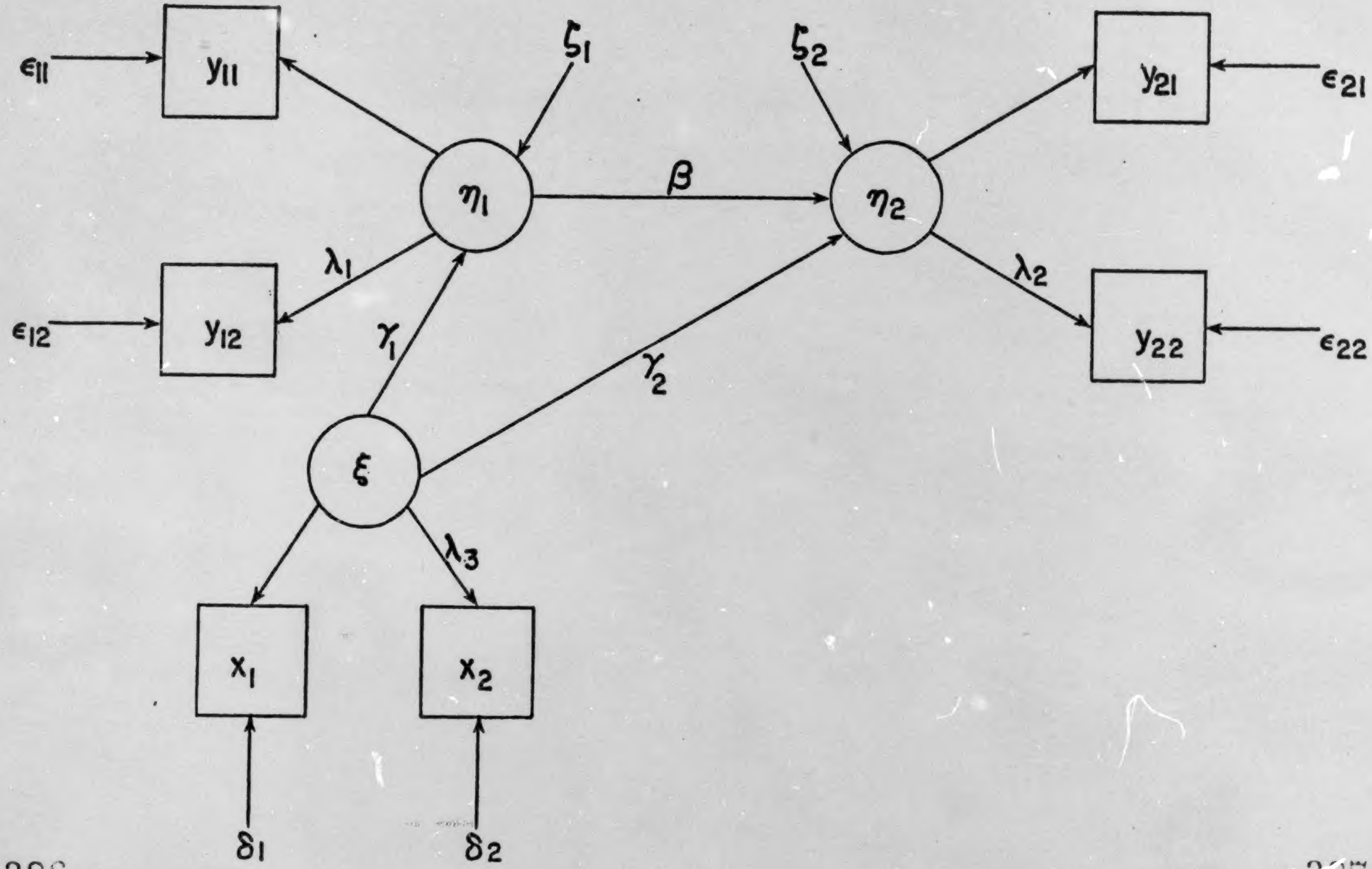

ε11
y11
ζ1
ζ2
y21
ε21
η1
β
η2
λ1
λ2
ε12
y12
γ1
γ2
y22
ε22
ξ
λ3
x1
x2
δ1
δ2

Figure 9 -- A two-wave, two-variable model with two congeneric background variables.

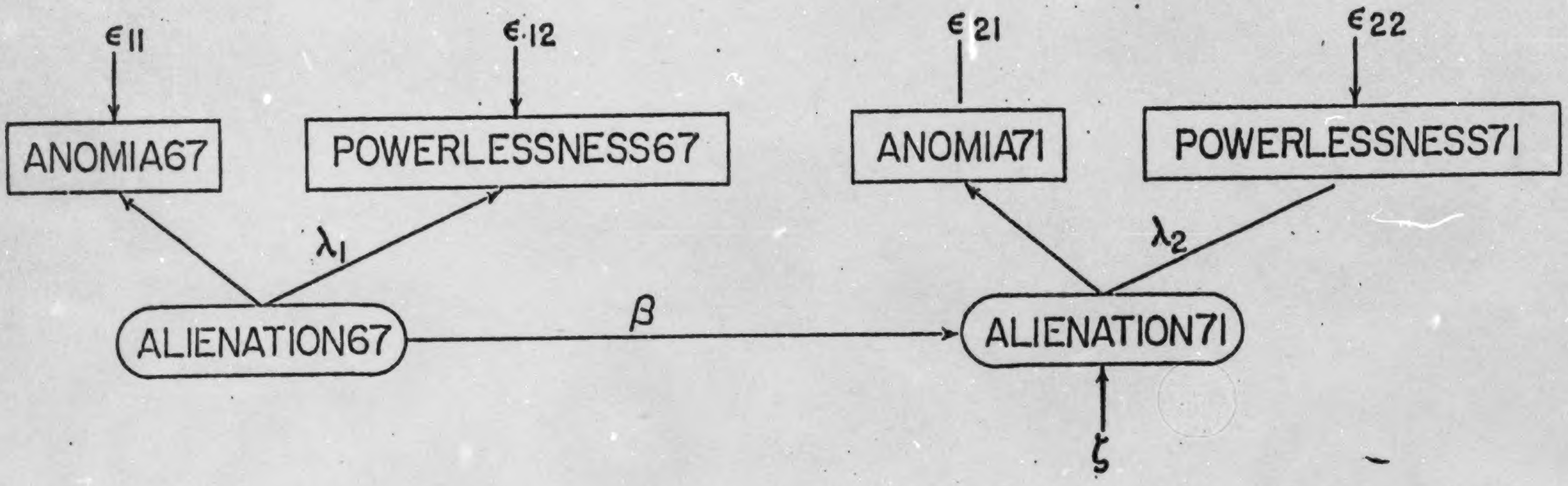

ε11
ε12
ε21
ε22
ANOMIA67
POWERLESSNESS67
ANOMIA71
POWERLESSNESS71
λ1
λ2
ALIENATION67
β
ALIENATION71
ζ

Figure 10a -- Model for study of stability of alienation.

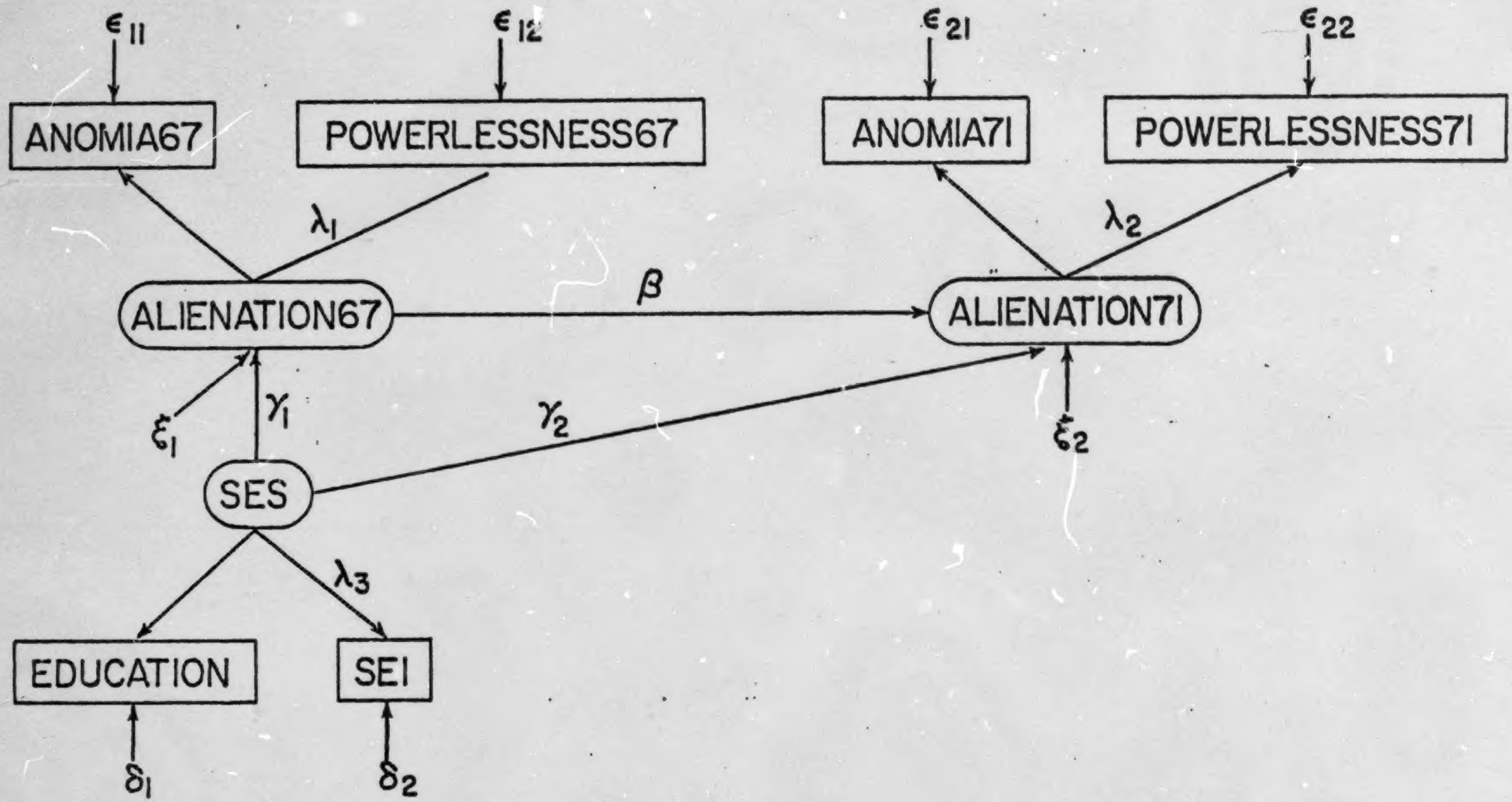

ϵ_{11}
ϵ_{12}
ϵ_{21}
ϵ_{22}
ANOMIA67
POWERLESSNESS67
ANOMIA71
POWERLESSNESS71
λ_1
λ_2
ALIENATION67
β
ALIENATION71
ξ_1
γ_1
γ_2
ξ_2
SES
λ_3
EDUCATION
SEI
δ_1
δ_2

Figure 10b -- Model for study of stability of alienation.

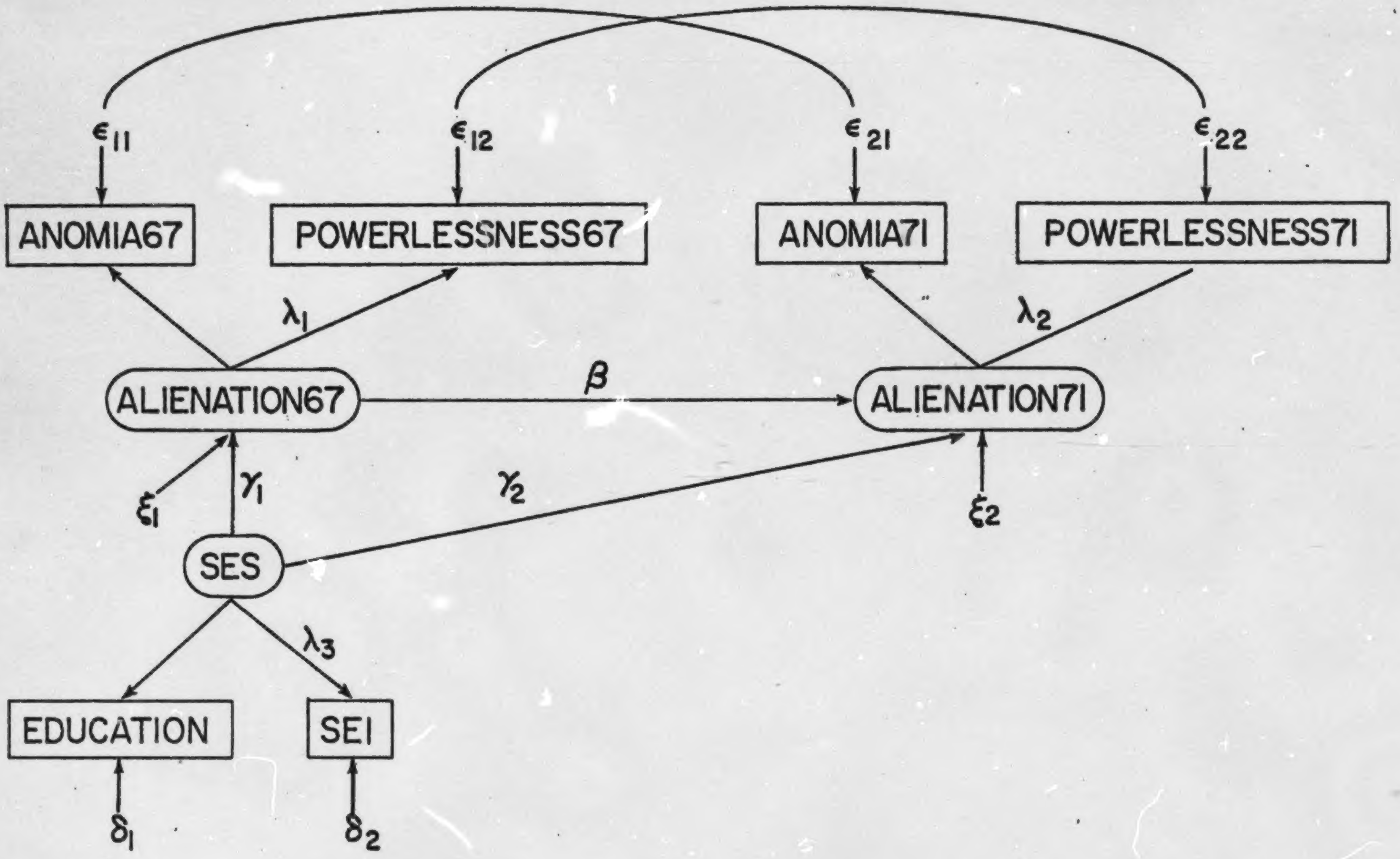

ϵ_{11}
ϵ_{12}
ϵ_{21}
ϵ_{22}
ANOMIA67
POWERLESSNESS67
ANOMIA71
POWERLESSNESS71
λ_1
λ_2
ALIENATION67
β
ALIENATION71
ξ_1
γ_1
γ_2
ξ_2
SES
λ_3
EDUCATION
SEI
δ_1
δ_2

Figure 10c -- Model for study of stability of alienation.

Table 2

Covariance Matrix for the Models of Figure 10 (N = 932)

Y_1	11.834					
Y_2	6.947	9.364				
Y_3	6.819	5.091	12.532			
Y_4	4.783	5.028	7.495	9.986		
X_1	- 3.839	- 3.889	- 3.841	- 3.625	9.610	
X_2	-21.899	-18.831	-21.748	-18.775	35.522	450.288

The maximum likelihood estimates of the parameters of the models are given in Table 3. The main aim of the Wheaton et al study was to estimate the stability of alienation over time, which is reflected in the parameter β, or in the correlation between alienation 71 and alienation 67. As can be seen from Table 3 we obtain an estimate of β which is biased upwards if we use a model that does not take SES into account. The influence of SES on Alienation at the two occasions is significant, see Model 10B. The coefficient for 1967, γ_1, is -0.614 with a standard error of 0.056 and for 1971, γ_2, it is -0.174 with a standard error equal to 0.054. The negative signs of the SES-coefficients γ_1 and γ_2 indicate that for high socioeconomic status the alienation is low and vice versa. However, the overall fit of the Model 10B is not acceptable; χ^2 with six degrees of freedom equals 71.544. Since the same scales are used on both occasions, it seems reasonable to assume that if the influence of the true score, i.e. Alienation, is removed from the measured variables, i.e. Anomia and Powerlessness, there might still be some correlation left between the same measures at the two occasions. Thus, the Model 10C is intuitively more plausible. As can be seen from Table 3 the inclusion of these error correlations results in a model with an acceptable overall fit.

D. An Analysis of Verbal and Quantitative Ability

For the second illustration we use some longitudinal data from a large growth study conducted at Educational Testing Service

----
Insert Table 3 about here

TABLE 3

Maximum Likelihood Estimates for the Models in Figure 10 A-C

The standard errors of the estimates are given within parenthesis.

	Model in Figure 10A	Model in Figure 10B	Model in Figure 10C
λ_1	0.815 (.040)	0.888 (.041)	0.979 (.062)
λ_2	0.847 (.042)	0.849 (.040)	0.922 (.059)
λ_3	- - -	5.331 (.430)	5.221 (.422)
β	0.789 (.044)	0.705 (.054)	0.607 (.051)
γ_1	- - -	-0.614 (.056)	-0.575 (.056)
γ_2	- - -	-0.174 (.054)	-0.227 (.052)
ψ_{11}	- - -	5.307 (.473)	4.847 (.468)
ψ_{22}	4.085 (.432)	3.742 (.388)	4.089 (.405)
ϕ	- - -	6.663 (.641)	6.803 (.650)
σ_{δ_1}	- - -	1.717 (.145)	1.675 (.151)
σ_{δ_2}	- - -	16.153 (.565)	16.273 (.558)
$\sigma_{\varepsilon_{11}}$	1.906 (.097)	2.004 (.086)	2.176 (.104)
$\sigma_{\varepsilon_{12}}$	1.865 (.077)	1.786 (.076)	1.602 (.126)
$\sigma_{\varepsilon_{21}}$	1.827 (.109)	1.923 (.097)	2.098 (.123)
$\sigma_{\varepsilon_{22}}$	1.969 (.077)	1.904 (.077)	1.754 (.124)
corr(ε_{11},ε_{21})	- - -	- - -	0.356 (.047)
corr(ε_{12},ε_{22})	- - -	- - -	0.121 (.082)
χ^2	61.155	71.544	4.770
d.f.	1	6	4

(Anderson & Maier, 1963; Hilton, 1969). In this study, a nationwide (U.S.A.) sample of fifth graders was tested in 1961 and then again in 1963, 1965 and 1967 as seventh, ninth and eleventh graders, respectively. The test scores include the verbal (SCATV) and quantitative (SCATQ) parts of the SCAT (Scholastic Aptitude Test) and achievement tests in mathematics (MATH), science (SCI), social studies (SS), reading (READ), listening (LIST), and writing (WRIT). The examinees for which complete data were available for all the grades 5, 7, 9 and 11 were divided into four groups according to sex and whether or not they were in the academic curriculum in grade 12. The four groups and their sample sizes are as follows:

Boys academic (BA): N = 373,
Boys non-academic (BNA): N = 249,
Girls academic (GA): N = 383,
Girls non-academic (GNA): N = 387.

Scores on each test have been scaled so that the unit of measurement is approximately the same at all occasions. All analyses reported here are based on information provided by the means, standard deviations and intercorrelations of the 32 variables (8 tests at 4 occasions) for the four groups.

In this example we use the six tests MATH, SCI, SS, READ, SCATV, SCATQ in grades 7 and 9 only and only the group GA. In later sections we use data from other grades and groups as well. Earlier studies (Jöreskog, 1970a) suggest that these tests measure two oblique factors which may reasonably be interpreted as a verbal (V) and a quantitative (Q) factor. We set up the model in Fig. 11, which represents a model for the measurement of change in verbal and quantitative ability between grades 7 and 9. Since there are no background variables in this model we may for estimation purposes treat the pretests as the independent variables. Hence we use the notation $\mathbf{x}$ for these. Note that the model includes the following features:

(i) On each occasion the factor pattern is postulated to be restricted in the following way. MATH and SCATQ are pure measures of Q. READ and SCATV are pure measures of V. SCI and SS and composite measures of V and Q. This implies that there are four zero loadings in both $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$. To fix the scales for V and Q we assume that they are measured in the same units as SCATV and SCATQ, respectively. This means that there is one fixed one in each column of $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$.

(ii) It is postulated that Q_7 affects Q_9 only and not V_9 and similarly for V_7. This means that there are two zero coefficients in $\boldsymbol{\Gamma}$. Furthermore, we postulate that the residuals ζ_1 and ζ_2 are uncorrelated, which means that, whatever remains in Q_9 and V_9 after Q_7 and V_7 are accounted for, is uncorrelated with everything else.

(iii) The errors or unqiue factors in $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are assumed to bb uncorrelated both within and between occasions.

Insert Figure 11 about here

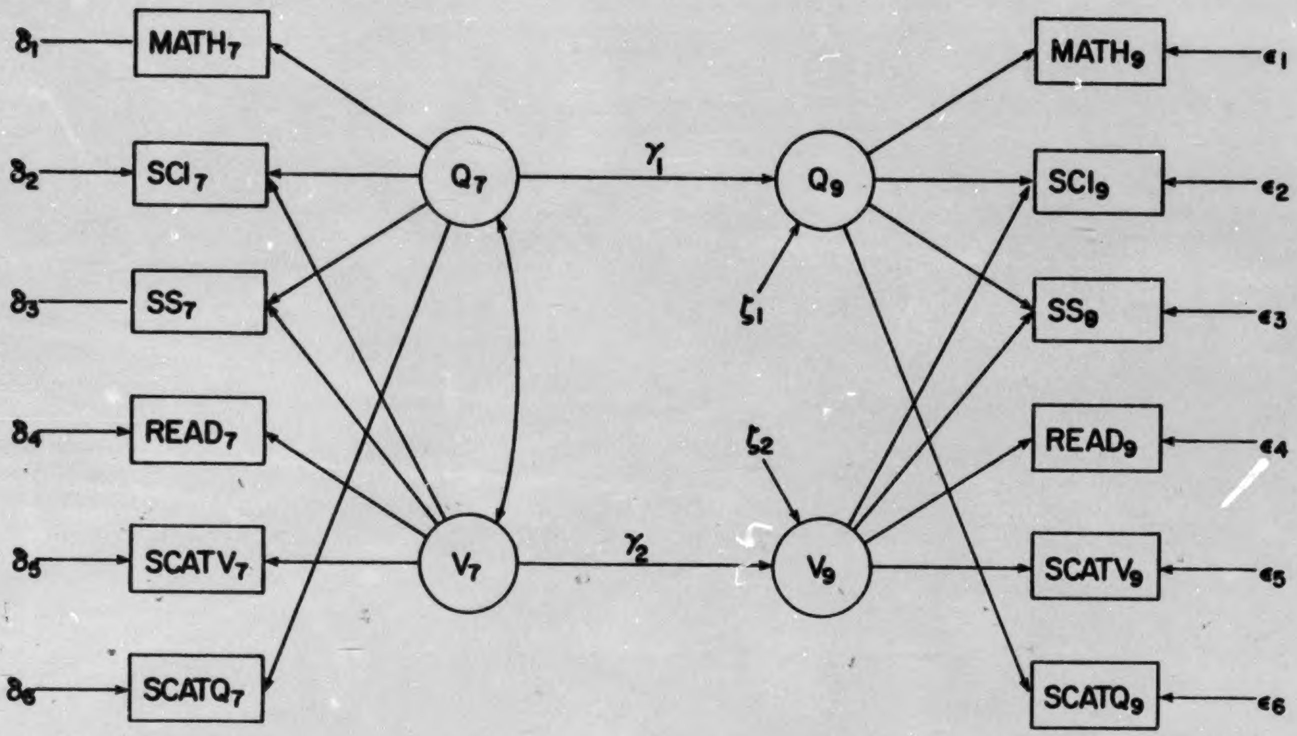

δ_1
MATH$_7$
δ_2
SCI$_7$
δ_3
SS$_7$
δ_4
READ$_7$
δ_5
SCATV$_7$
δ_6
SCATQ$_7$
Q$_7$
V$_7$
γ_1
γ_2
Q$_9$
V$_9$
ζ_1
ζ_2
MATH$_9$
SCI$_9$
SS$_9$
READ$_9$
SCATV$_9$
SCATQ$_9$
ϵ_1
ϵ_2
ϵ_3
ϵ_4
ϵ_5
ϵ_6

Figure 11 -- Model for the measurement of change in verbal and quantitative ability between grades 7 and 9.

The maximum likelihood estimates are given in Table 4. The rather low loadings of SCI and SS on Q at both occasions may seem a little surprising. However, an inspection of the items in tests SCI and SS reveals that these are mostly verbal problems concerned with logical reasoning in contrast to the items in SCATQ which are mostly numerical items measuring the ability to work with numbers. The small residual variance 1.85 of ζ_2 means that V_9 can be predicted almost perfectly from V_7. This is not quite so for Q since we here have a residual variance of 18.49. However, this may be due to the more rapid increase in variance of Q from grade 7 to 9, which is manifested in the increase in variances which is 143.54 - 103.87 = 39.67 for Q and 117.15 - 115.41 = 1.74 for V.

There is a reason not to look at each number in Table 4 too seriously and this is the poor overall fit of the model as evidenced by the χ^2-value of 217.79 with 47 degrees of freedom. We shall therefore investigate the reason for this poor fit and demonstrate that LISREL may be used not only to assess or measure the goodness of fit of a model but also to detect the parts of the model where the fit is poor. Taking the more fundamental assumptions of linearity and multinormality for granted, lack of fit of the model in Fig. 11 may be due to one or more of the postulates (i), (ii) or (iii) not being reasonable. We shall therefore investigate each of these separately.

To investigate (i), we set up a factor analysis of the pre- and posttests separately assuming the postulated two-factor structure.

Insert Table 4 about here

TABLE 4

Maximum Likelihood Estimates (LISREL) for the Model of Figure 11

Group: GA (Girls Academic), N = 383

$$\underset{\sim}{\hat{\Lambda}}_x = \begin{array}{cc} \begin{array}{cc} Q_7 & \quad V_7 \end{array} & \begin{array}{cc} \phantom{\mathrm{SCATQ}_7} & \theta_\delta \end{array} \\ \begin{bmatrix} 0.97 & 0.^* \\ 0.20 & 0.52 \\ 0.25 & 0.84 \\ 0.^* & 1.21 \\ 0.^* & 1.^* \\ 1.^* & 0.^* \end{bmatrix} & \begin{array}{ll} \mathrm{MATH}_7 & 5.68 \\ \mathrm{SCI}_7 & 5.49 \\ \mathrm{SS}_7 & 6.61 \\ \mathrm{READ}_7 & 6.80 \\ \mathrm{SCATV}_7 & 4.44 \\ \mathrm{SCATQ}_7 & 7.10 \end{array} \end{array}$$

$$\underset{\sim}{\hat{\Lambda}}_x = \begin{array}{cc} \begin{array}{cc} Q_9 & \quad V_9 \end{array} & \begin{array}{cc} \phantom{\mathrm{SCATQ}_9} & \theta_\xi \end{array} \\ \begin{bmatrix} 0.88 & 0.^* \\ 0.24 & 0.64 \\ 0.36 & 0.69 \\ 0.^* & 0.95 \\ 0.^* & 1.^* \\ 1.^* & 0.^* \end{bmatrix} & \begin{array}{ll} \mathrm{MATH}_9 & 4.80 \\ \mathrm{SCI}_9 & 6.57 \\ \mathrm{SS}_9 & 7.24 \\ \mathrm{READ}_9 & 6.49 \\ \mathrm{SCATV}_9 & 4.47 \\ \mathrm{SCATQ}_9 & 8.23 \end{array} \end{array}$$

$$\underset{\sim}{\hat{\Gamma}} = \begin{array}{cc} \begin{array}{cc} Q_7 & \quad V_7 \end{array} & \\ \begin{pmatrix} 1.10 & 0.^* \\ 0.^* & 1.00 \end{pmatrix} & \begin{array}{l} Q_9 \\ V_9 \end{array} \end{array}$$

$$\underset{\sim}{\hat{\Phi}} = \begin{array}{cc} \begin{array}{cc} Q_7 & \quad V_7 \end{array} & \\ \begin{pmatrix} 103.87 & 92.58 \\ 92.58 & 115.41 \end{pmatrix} & \begin{array}{l} Q_7 \\ V_7 \end{array} \end{array}$$

$$\underset{\sim}{\hat{\Psi}} = \begin{array}{cc} \begin{array}{cc} \zeta_1 & \quad \zeta_2 \end{array} & \\ \begin{pmatrix} 18.49 & 0.^* \\ 0.^* & 1.85 \end{pmatrix} & \begin{array}{l} \zeta_1 \\ \zeta_2 \end{array} \end{array}$$

$$\hat{\Omega} = \begin{array}{cc} \begin{array}{cc} Q_9 & \quad V_9 \end{array} & \\ \begin{pmatrix} 143.54 & 101.54 \\ 101.54 & 117.15 \end{pmatrix} & \begin{array}{l} Q_9 \\ V_9 \end{array} \end{array}$$

$\chi^2 =$ 217.79 with d.f. = 47

* the value of this parameter was specified by the model.

This gives $\chi^2 = 17.64$ for the pretests and $\chi^2 = 2.62$ for the posttests both with 10 degrees of freedom. Although the fit is not quite acceptable in grade 7 we take the postulated factor structure to hold both for the pre- and posttests. So we must continue to look for lack of fit due to (ii) or (iiii).

The postulate (ii) is concerned with the interrelationships between the four factors Q_7, V_7, Q_9, and V_9. The most general assumption is that these four factors are freely intercorrelated and this is equivalent to a LISREL model with all four coefficients in $\underset{\sim}{\Gamma}$ free and with $\underset{\sim}{\Psi}$ free as a full symmetric matrix. Hence, it is clear that the assumptions made in (ii) is the intersection of the two hypotheses "$\underset{\sim}{\Gamma}$ is diagonal" and "$\underset{\sim}{\Psi}$ is diagonal". It is therefore useful to test each of the four possible hypotheses. The results of these analyses may be presented in a 2 x 2 table as in Table 5. The row marginals of the table represent χ^2-values with one degrees of freedom for testing the hypothesis that $\underset{\sim}{\Psi}$ is diagonal. It is seen that this hypothesis may be rejected. The column marginals represent χ^2-values with two degrees of freedom for testing the hypothesis that $\underset{\sim}{\Gamma}$ is diagonal. This hypothesis seems quite reasonable. From these analyses it is clear that "$\underset{\sim}{\Gamma}$ diagonal and $\underset{\sim}{\Psi}$ free" is the most reasonable assumption to retain. The overall fit of this model is $\chi^2 = 196.4$ with 46 degrees of freedom. Since this is still too large we must continue to investigate (iii).

Insert Table 5 about here

TABLE 5

Test of Assumptions (ii) for the Model in Figure 11

Group: GA (Girls Academic), N = 383

	$\underset{\sim}{\Psi}$ diagonal	$\underset{\sim}{\Psi}$ free	
$\underset{\sim}{\Gamma}$ diagonal	$\chi^2_{47} = 217.8$	$\chi^2_{46} = 196.4$	$\chi^2_{1} = 21.4$
$\underset{\sim}{\Gamma}$ free	$\chi^2_{45} = 216.8$	$\chi^2_{44} = 193.7$	$\chi^2_{1} = 23.1$
	$\chi^2_{2} = 1.0$	$\chi^2_{2} = 2.7$	

The assumption in (iii) is that the unique factors in $\underset{\sim}{\delta}$ and $\underset{\sim}{\varepsilon}$ are uncorrelated both within and between sets. That they are uncorrelated within sets should not be questioned since we have already found that the postulated factor analysis model holds for both pre- and posttest. That they are uncorrelated between sets, however, is more questionable because of specific factors in each test. This means that the unique factors for corresponding tests should be allowed to correlate. To account for such correlations, Jöreskog (1970a) introduced so called test-specific factors, i.e. factors which do not contribute to correlations between tests within occasions but between the same tests at different occasions. In this case, when there are only two occasions, it is not possible to define (identify) test specific factors but we can merely introduce correlations between unique factors for corresponding pre- and posttests.

The model in Fig. 11 is therefore modified as in Fig. 12. This revised model can also be estimated with the LISREL program. The analysis of the revised model gives the results shown in Table 6 which also gives standard errors of the estimated parameters. It is seen that all the estimated parameters are significantly different from zero. The test of overall goodness of fit gives $\chi^2 = 65.63$ with 40 degrees of freedom. This represents a reasonably good fit of the model to the data. An approximate test of the hypotheses that the unique factors are uncorrelated between occasions is

----
Insert Figure 12 and Table 6 about here

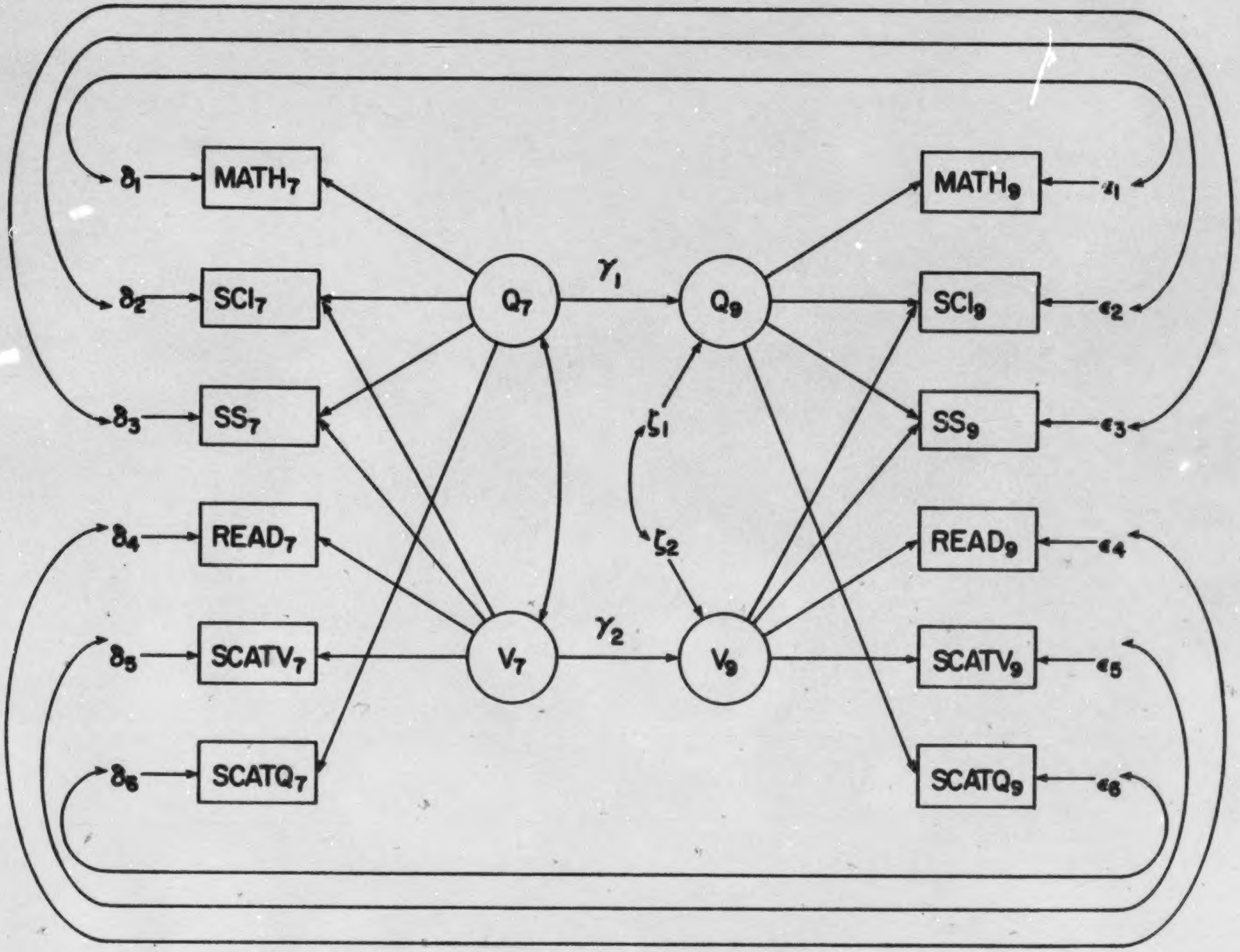

δ_1
MATH$_7$
δ_2
SCI$_7$
δ_3
SS$_7$
δ_4
READ$_7$
δ_5
SCATV$_7$
δ_6
SCATQ$_7$
Q$_7$
V$_7$
γ_1
γ_2
ζ_1
ζ_2
Q$_9$
V$_9$
MATH$_9$
SCI$_9$
SS$_9$
READ$_9$
SCATV$_9$
SCATQ$_9$
ϵ_1
ϵ_2
ϵ_3
ϵ_4
ϵ_5
ϵ_6

Figure 12 -- Revised model for the measurement of change in verbal and quantitative ability between grades 7 and 9.

TABLE 6

Maximum Likelihood Estimates (LISREL) for the Model of Figure 12

Standard Errors in Parenthesis

Group: GA (Girls Academic), N = 383

$$\hat{\Lambda}_x = \begin{array}{lll} Q_7 & V_7 & \\ 1.01\ (0.05) & 0.^* & \mathrm{MATH}_7 \\ 0.13\ (0.07) & 0.60\ (0.07) & \mathrm{SCI}_7 \\ 0.12\ (0.09) & 0.98\ (0.09) & \mathrm{SS}_7 \\ 0.^* & 1.24\ (0.05) & \mathrm{READ}_7 \\ 0.^* & 1.^* & \mathrm{SCATV}_7 \\ 1.^* & 0.^* & \mathrm{SCATQ}_7 \end{array}$$

$$\hat{\Lambda}_y = \begin{array}{lll} Q_9 & V_9 & \\ 0.93\ (0.05) & 0.^* & \mathrm{MATH}_9 \\ 0.13\ (0.07) & 0.77\ (0.08) & \mathrm{SCI}_9 \\ 0.25\ (0.08) & 0.82\ (0.08) & \mathrm{SS}_9 \\ 0.^* & 0.98\ (0.04) & \mathrm{READ}_9 \\ 0.^* & 1.^* & \mathrm{SCATV}_9 \\ 1.^* & 0.^* & \mathrm{SCATQ}_9 \end{array}$$

$$\Gamma = \begin{array}{lll} Q_7 & V_7 & \\ 1.06\ (0.05) & 0.^* & Q_9 \\ 0.^* & 0.98\ (0.03) & V_9 \end{array}$$

$$\hat{\Phi} = \begin{array}{lll} Q_7 & V_7 & \\ 100.57\ (10.86) & 90.53\ (8.46) & Q_7 \\ 90.53\ (8.46) & 110.45\ (9.74) & V_7 \end{array}$$

$$\hat{\Psi} = \begin{array}{lll} \S_1 & \S_2 & \\ 22.63\ (4.41) & 8.42\ (1.73) & \S_1 \\ 8.42\ (1.73) & 6.94\ (1.58) & \S_2 \end{array}$$

$$\hat{\Omega} = \begin{array}{lll} Q_9 & V_9 & \\ 136.52 & 102.62 & Q_9 \\ 102.62 & 112.53 & V_9 \end{array}$$

χ^2 = 65.63 with d.f. = 40

* the value of this parameter was specified by the model

obtained as $\chi^2 = 196.4 - 65.6 = 130.8$ with 6 degrees of freedom so that it is clear that this hypothesis is quite unreasonable. The variances, covariances and correlations of the unique factors are given in Table 7. A comparison of the covariances with their standard errors reveals that all covariances except possibly the one between δ_1 and ε_1 is significantly non-zero.

E. Comparison of change between groups

In many longitudinal studies both pretests and posttests are administered to several groups of individuals and one is interested in comparing the change in various quantities between the different groups. Such groups may be, for example, groups having different socio-economic background, groups having obtained different amounts of schooling or training either prior to the pretest occasion or between the two occasions or groups having obtained different treatments between the two occasions. When we have several groups it is natural to assume that the distributions of the latent variables are different for the different groups. Sörbom (1974) has developed a model in which the mean vector as well as the covariance matrix of the latent variables may vary from group to group. The structural equations will therefore be different for different groups. On the other hand, the matrix $\underset{\sim}{\Lambda}$, which describes the relationships between the observed test scores and the latent variables, is considered an attribute of the observed variables and is therefore assumed to be the same for all groups.

It is assumed that observations from different groups are independent. For a "random" examinee from group g we write his

---

Insert Table 7 about here

TABLE 7

Variances, Covariances and Correlations for the Unique Factors in Table 6 (Figure 12)

Standard Errors in Parenthesis

i	Var (δ_i)	Var (ε_i)	Cov ($\delta_i \varepsilon_i$)	Corr ($\delta_i \varepsilon_i$)
1	27.75 (3.78)	17.67 (3.88)	-3.47 (2.78)	0.157
2	29.59 (2.37)	41.15 (3.38)	9.60 (2.10)	0.275
3	40.27 (3.60)	50.89 (4.21)	6.15 (2.82)	0.136
4	44.21 (4.25)	40.34 (3.59)	7.52 (2.89)	0.178
5	24.37 (2.47)	24.84 (2.60)	12.04 (2.05)	0.489
6	54.28 (4.87)	74.53 (6.73)	22.84 (4.40)	0.359

observed scores, using the notation in (9)

$$\underset{\sim}{z}_g = \underset{\sim}{\nu} + \underset{\sim}{\Lambda}\underset{\sim}{f}_g + \underset{\sim}{e}_g \quad . \tag{33}$$

Note that both $\underset{\sim}{\nu}$ and $\underset{\sim}{\Lambda}$ are the same for all groups. The constant vector $\underset{\sim}{\nu}$ represents the origin or level of the tests in the sense that when $\underset{\sim}{f}_g = \underset{\sim}{0}$ then $E(\underset{\sim}{z}_g) = \underset{\sim}{\nu}$ for all groups. This is considered an attribute of the tests and the scoring procedure. Let the mean vector of $\underset{\sim}{f}_g$ be denoted $\underset{\sim}{\theta}_g$. Then the mean vector $\underset{\sim}{\mu}_g$ of $\underset{\sim}{z}_g$ is (c.f. equation (10))

$$\underset{\sim}{\mu}_g = \underset{\sim}{\nu} + \underset{\sim}{\Lambda}\underset{\sim}{\theta}_g \quad . \tag{34}$$

and the covariance matrix is (c.f. equation (11))

$$\underset{\sim}{\Sigma}_g = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}_g\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}_g \tag{35}$$

where $\underset{\sim}{\Phi}_g$ is the covariance matrix of $\underset{\sim}{f}_g$ and $\underset{\sim}{\Psi}_g$ the covariance matrix of $\underset{\sim}{e}_g$.

There are two fundamental indeterminacies in (34) and (35). Every factor in $\underset{\sim}{f}_g$ may be subjected to an arbitrary linear transformation which may be different for different factors but the same for all individuals in all groups. The effect of such transformations may be compensated for by adding a constant vector to $\underset{\sim}{\nu}$ and by a scaling of the columns of $\underset{\sim}{\Lambda}$, in such a way that both $\underset{\sim}{\mu}_g$ and $\underset{\sim}{\Sigma}_g$ are unchanged for all groups. This indeterminacy means that both the origin and the scale for the factors are arbitrary. These may therefore be chosen arbitrarily, but must be the same for all groups. It is convenient to fix the origins and the scales by choosing the vector $\underset{\sim}{\theta}$ equal to $\underset{\sim}{0}$ for one group and by choosing a one in each column of $\underset{\sim}{\Lambda}$.

Models of this kind may be estimated by means of the COFAMM program described in II. This gives maximum likelihood estimates of the common $\underset{\sim}{\nu}$ and $\underset{\sim}{\Lambda}$ and of the mean vector $\underset{\sim}{\theta}_g$ and covariance matrix $\underset{\sim}{\Phi}_g$ as well as the covariance matrix $\underset{\sim}{\Psi}_g$ of the unique factors for each group. One may postulate almost any pattern in $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}_g$ and $\underset{\sim}{\Psi}_g$ and any degree of invariance between groups. For example, one may postulate that $\underset{\sim}{\Psi}_g$ and some part of $\underset{\sim}{\Phi}_g$ are invariant over groups.

F. Comparison of Change in Verbal Ability between Groups

To illustrate the method of the preceding subsection we make use of the data introduced in section IV.D. This time we use the data for all the four groups but we use a somewhat simpler model than that of Figures 11 and 12. We shall use scores on the reading and writing achievement tests in grades 7 and 9 only. The model is shown in Fig. 13. Here we are mainly concerned with the comparison of the differences in mean changes and in the regression lines of V_9 on V_7.

Insert Figure 13 about here

The regression of V_9 on V_7 in group g is

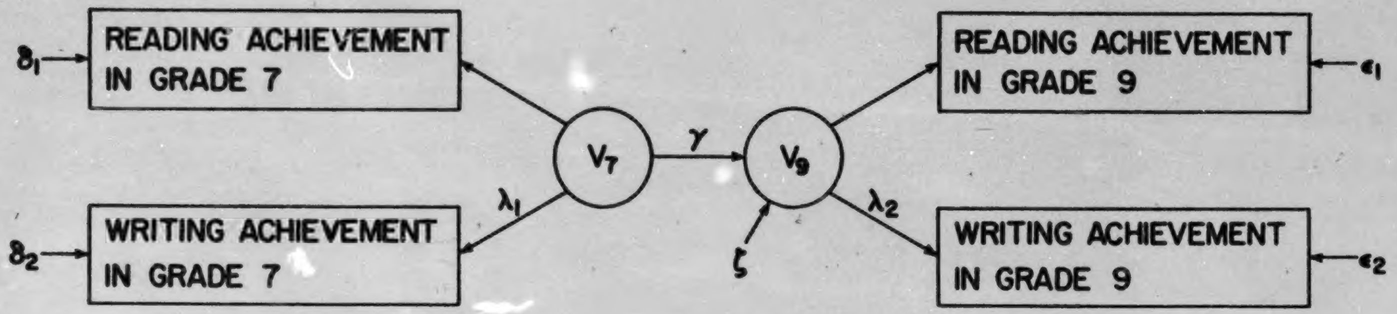

δ₁
READING ACHIEVEMENT IN GRADE 7
δ₂
WRITING ACHIEVEMENT IN GRADE 7
λ₁
V₇
γ
V₉
ζ
λ₂
READING ACHIEVEMENT IN GRADE 9
ε₁
WRITING ACHIEVEMENT IN GRADE 9
ε₂

Figure 13 -- Model for comparison of change in verbal ability between groups.

$$V_9 = \alpha_g + \gamma_g V_7 + \zeta ,$$

where

$$\gamma_g = \sigma_{\eta\xi g}/\sigma^2_{\xi g},$$

$$\alpha_g = \theta_{\eta g} - \gamma_g \theta_{\xi g} .$$

The maximum likelihood solution is given in Table 8. If one takes the intercepts $\hat{\alpha}_g$ as relative measures of change, remembering that the scale is chosen such that α_g is zero for group BA , one finds that group GA has increased their verbal ability most followed by groups BA , GNA and BNA in that order. However, this is not the whole story. For since the slope of the regression lines also differ between groups one should take this also into account when interpreting the data. Probably the best way of looking at the results is to use the estimates $\hat{\underset{\sim}{\theta}}_g$ and $\hat{\underset{\sim}{\Phi}}_g$ to draw contour ellipses for each group as in Fig. 14. With this kind of plot one can fix a given true pretest score and find the likely range of true posttest score for the various groups. For example, at $\xi = -15$, approximate 95% confidence intervals for η are

GA: $-17.32 \leq \eta \leq -4.63$,

PA: $-22.40 \leq \eta \leq -1.038$,

GNA: $-26.36 \leq \eta \leq -0.87$,

BNA: $-29.60 \leq \eta \leq -2.93$.

At $\xi = 10$ these confidence intervals show a different pattern:

GA: $0.18 \leq \eta \leq 16.34$

BA: $-3.46 \leq \eta \leq 19.09$

GNA: $-3.62 \leq \eta \leq 19.18$

BNA: $3.02 \leq \eta \leq 13.46$.

Insert Figure 14 and Table 8 about here

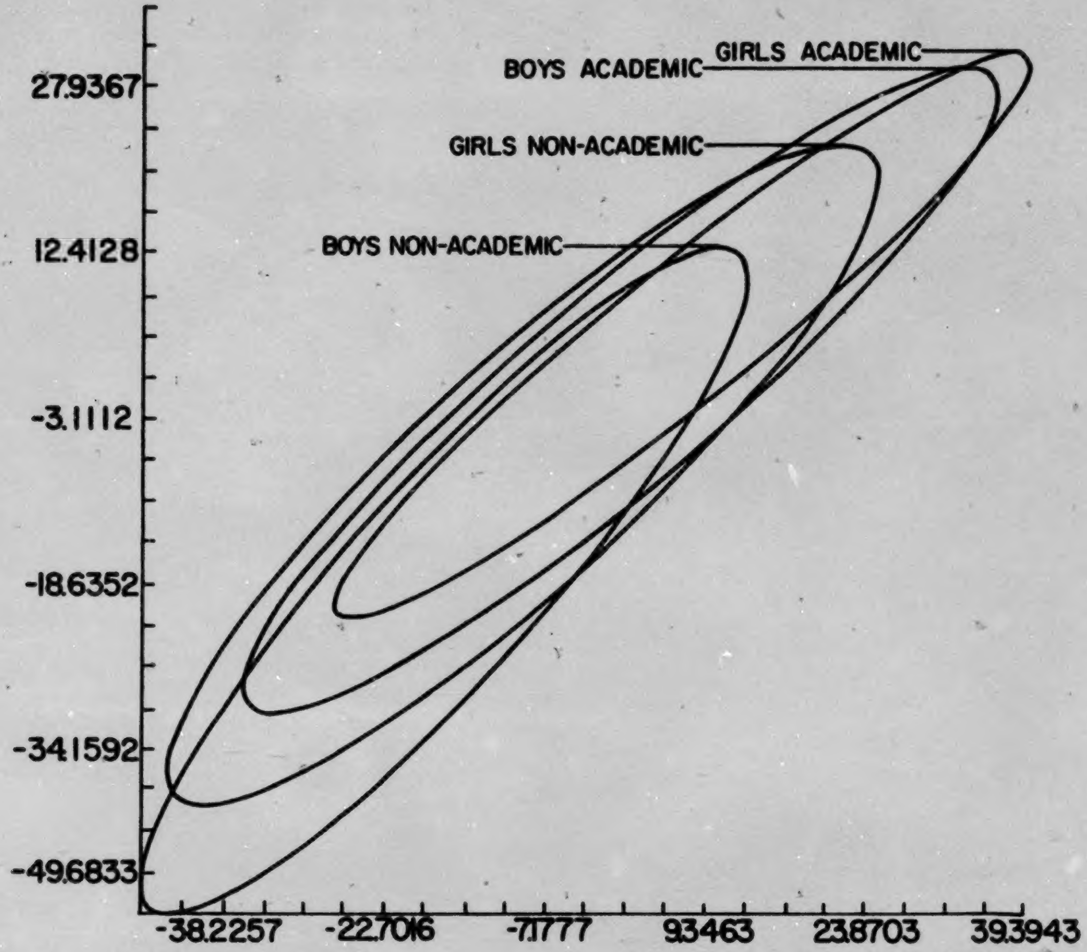


Figure 14 -- ETS growth study: Verbal ability, grade 7 - 9. 95 per cent regions.

TABLE 8

Maximum Likelihood Estimates for the Model of Figure 13

Simultaneous Analysis for all Four Groups

Groups:` BA (Boys academic) N = 373 GA (Girls academic) N = 383

BNA (Boys non-academic) N = 249 GNA (Girls non-academic) N = 387

$$\underset{\sim}{\Lambda} = \begin{bmatrix} 1.00^* & 0.00^* \\ 0.95 & 0.00^* \\ 0.00^* & 1.00^* \\ 0.00^* & 1.12 \end{bmatrix} \qquad \underset{\sim}{\nu} = \begin{bmatrix} 274.93 \\ 269.76 \\ 286.83 \\ 283.35 \end{bmatrix}$$

g	$\hat{\sigma}_{\xi g}$	$\hat{\sigma}_{\eta\xi g}$	$\hat{\sigma}_{\eta g}$	$\hat{\gamma}_g$
BA	219.46	171.48	156.97	0.78
BNA	142.35	139.53	166.55	0.98
GA	186.65	143.57	121.49	0.77
GNA	195.17	160.24	163.38	0.82

g	$\theta_{\xi g}$	$\theta_{\eta g}$	α_g
BA	0.00*	0.00*	0.00*
BNA	-16.881	-18.108	-1.56
GA	5.949	5.140	0.56
GNA	-9.298	-8.061	-0.43

V. Multi-Wave Models

A. Multi-Wave, One-Variable Models

Suppose one fallible measure y is administered repeatedly to the same group of people. An appropriate model for this situation is shown in Fig. 15 in the case of four occasions. In the following

----

Insert Figure 15 about here

----

we discuss all models in terms of four occasions, the generalization to an arbitrary number of occasions will be obvious at all stages. Such models have been termed simplex models by Guttman (1954) to designate the typical pattern of intercorrelations they give rise to. Anderson (1960) formulated this model in terms of various stochastic processes and treated the identification problem and Jöreskog (1970b) treated the estimation problem. An application to the measurement of academic growth has been given by Werts, Linn and Jöreskog (1977) and applications to sociological panel analysis have been discussed by Heise (1969), Wiley and Wiley (1970) and Werts, Jöreskog and Linn (1971).

The unit of measurement in the factors η_i may be chosen to be the same as in y_i , i = 1,2,3,4. The equations defining the model are then, taking all variables as deviations from their mean,

$$y_i = \eta_i + \varepsilon_i \, , \; i = 1,2,3,4 \, , \qquad (36)$$

$$\eta_i = \beta_i \eta_{i-1} + \zeta_i \, , \; i = 2,3,4, \qquad (37)$$

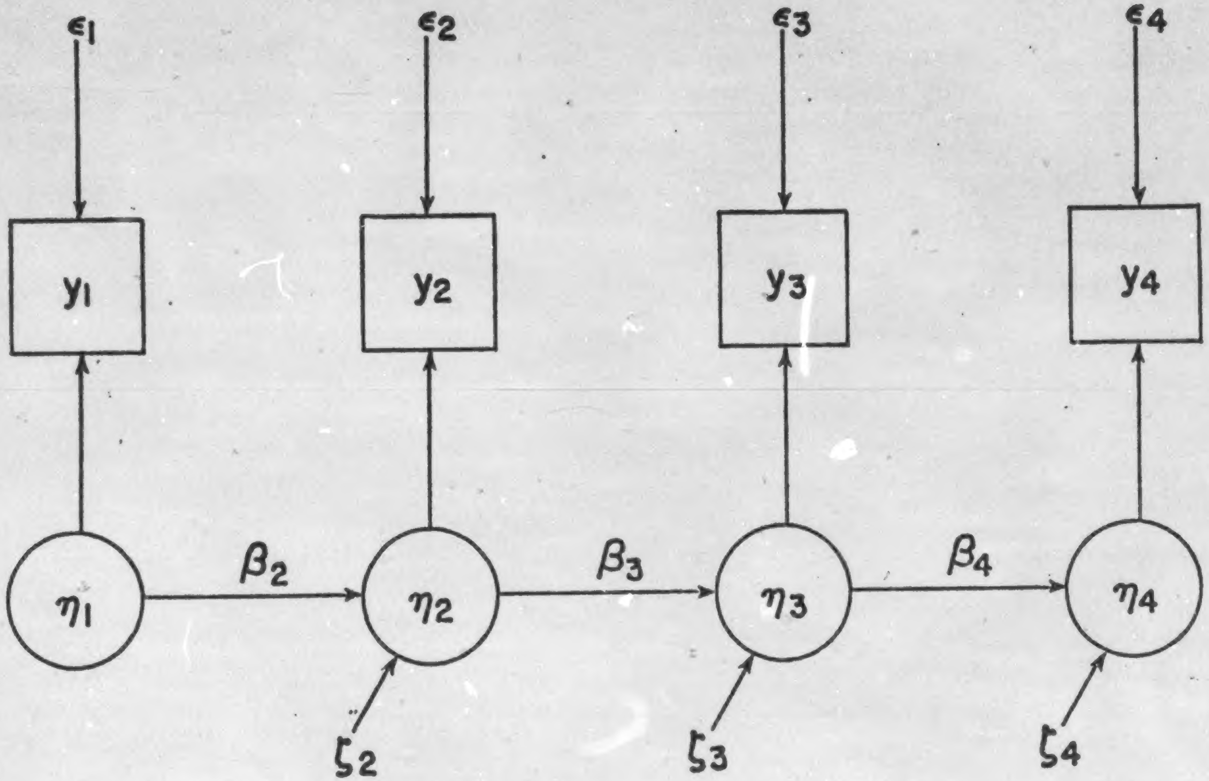

ϵ_1
ϵ_2
ϵ_3
ϵ_4
y_1
y_2
y_3
y_4
β_2
β_3
β_4
η_1
η_2
η_3
η_4
ζ_2
ζ_3
ζ_4

Figure 15 -- A simplex model.

where the ε_i are uncorrelated among themselves and uncorrelated with all the η_i and where ζ_{i+1} is uncorrelated with η_i, $i=1,2,3$. The parameters of the model are $\phi_i = \mathrm{Var}(\eta_i)$, $\theta_{ii} = \mathrm{Var}(\varepsilon_i)$, $i = 1,2,3,4$ and $\beta_2, \beta_3, \beta_4$. (We use the symbol ϕ here to denote the variances of the dependent variables. Since there are no independent variables there should be no confusion.) The residual variance $\mathrm{Var}(\zeta_{i+1})$ is a function of ϕ_{i+1}, ϕ_i and β_{i+1}, namely $\mathrm{Var}(\zeta_{i+1}) = \phi_{i+1} - \beta_{i+1}^2\phi_i$, $i=1,2,3$. The covariance matrix of y_1, y_2, y_3 and y_4 is

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \phi_1+\theta_{11} & & & \\ \beta_2\phi_1 & \phi_2+\theta_{22} & & \\ \beta_2\beta_3\phi_1 & \beta_3\phi_2 & \phi_3+\theta_{33} & \\ \beta_2\beta_3\beta_4\phi_1 & \beta_3\beta_4\phi_2 & \beta_4\phi_3 & \phi_4+\theta_{44} \end{bmatrix}. \quad (38)$$

It is seen from (38) that although the product $\beta_2\phi_1 = \sigma_{21}$ is identified, β_2 and ϕ_1 are not separately identified. The product $\beta_2\phi_1$ is involved in the off-diagonal elements in the first column (and row) only. We can multiply β_2 by a constant and divide ϕ_1 by the same constant without changing the product. The change induced by ϕ_1 in σ_{11} can be absorbed in θ_{11} in such a way that σ_{11} remains unchanged. Hence $\theta_{11} = \mathrm{Var}(\varepsilon_1)$ is not identified. For η_2 and η_3 we have

$$\phi_2 = \frac{\sigma_{32}\sigma_{21}}{\sigma_{31}} ,$$

$$\phi_3 = \frac{\sigma_{43}\sigma_{32}}{\sigma_{42}} ,$$

so that ϕ_2 and ϕ_3, and hence also θ_{22} and θ_{33}, are identified. With ϕ_2 and ϕ_3 identified, β_3 and β_4 are identified by σ_{32} and σ_{43}. The middle coefficient β_3 is overidentified since

$$\beta_3\phi_2 = \frac{\sigma_{31}\sigma_{42}}{\sigma_{41}} = \sigma_{32}.$$

Since both ϕ_4 and θ_{44} are involved in σ_{44} only, these are not identified but their sum σ_{44} is.

This analysis of the identification problem shows that for the "inner" variables y_2 and y_3, ϕ_2, ϕ_3, θ_{22}, θ_{33} and β_3 are identified, whereas there is an indeterminacy associated with each of the "outer" variables y_1 and y_4. To eliminate these indeterminancies one of the parameters ϕ_1, θ_{11} and β_2, must be specified and one of the parameters ϕ_4 and θ_{44} must also be specified. Hence there are only nine independent parameters and the model has one degree of freedom. In the general case of $T \geq 4$ occasions there will be $3T - 3$ free parameters and the degrees of freedom is $(1/2)T(T+1) - (3T-3)$.

The estimation problem associated with the simplex model is a straight-forward application of the LISREL program using the option of "no - x". The LISREL equations are

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{pmatrix} 0 \\ \varepsilon_2 \\ \varepsilon_3 \\ 0 \end{pmatrix} \tag{39}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 \\ 0 & 0 & -\beta_4 & 1 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix} \qquad (40)$$

In (39) we have taken $\varepsilon_1 = \varepsilon_4 = 0$ to eliminate the indeterminacies and in (40) we have defined ζ_1 as η_1. In LISREL it is inconvenient to treat $\phi_i = \mathrm{Var}(\eta_i)$, $i = 1,2,3,4$ as free parameters, so instead of $\phi_i = \mathrm{Var}(\eta_i)$, $i = 1,2,3,4$ we take $\psi_i = \mathrm{Var}(\zeta_i)$, $i = 1,2,3,4$ as free parameters. It is easily realized that the ϕ_i and the ψ_i, $i = 1,2,3,4$ are in a one-to-one correspondence. So the parameter matrices in LISREL are

$$\underset{\sim}{\Lambda}_y = \underset{\sim}{I} \ , \quad \underset{\sim}{B} \text{ as in (40)},$$

$$\underset{\sim}{\Psi} = \mathrm{diag}(\psi_1, \psi_2, \psi_3, \psi_4)$$

and

$$\underset{\sim}{\Theta}_\varepsilon = \mathrm{diag}(0, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}, 0).$$

B. Multi-Wave, Two-Variable Models

The direct generalization of the model in Fig. 5 to the case of four occasions is shown in Fig. 16 .

Insert Figure 16 about here

With $\underset{\sim}{x}' = (x_1, x_2, x_3, x_4)$, $\underset{\sim}{y}' = (y_1, y_2, y_3, y_4)$, the model is

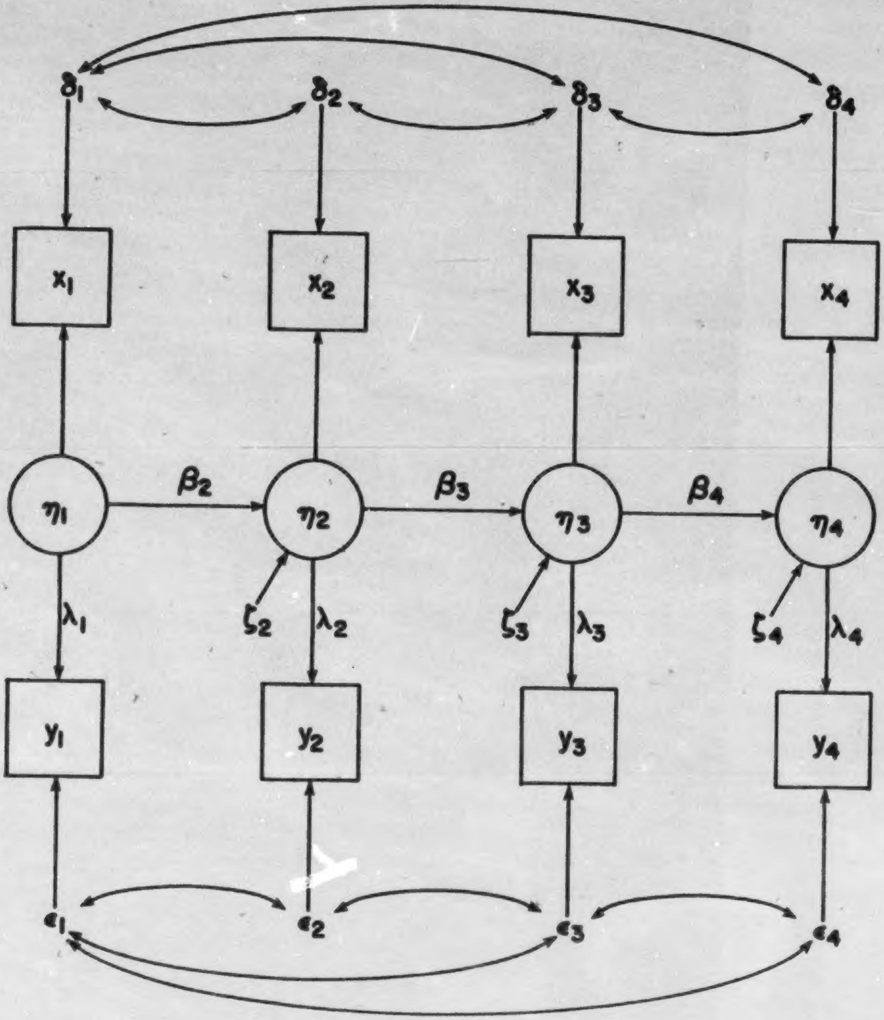

δ1
δ2
δ3
δ4
x1
x2
x3
x4
β2
β3
β4
η1
η2
η3
η4
λ1
ζ2
λ2
ζ3
λ3
ζ4
λ4
y1
y2
y3
y4
ε1
ε2
ε3
ε4

Figure 16 -- A four-wave, two-variable model with correlated errors.

$$\mathbf{x} = \boldsymbol{\eta} + \boldsymbol{\delta}, \tag{41}$$

$$\mathbf{y} = \mathbf{D}_\lambda \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{42}$$

where $\mathbf{D}_\lambda = \text{diag}\ (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. The covariance matrix of $\mathbf{z} = (\mathbf{x}' , \mathbf{y}')'$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} ,$$

with

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Phi} + \boldsymbol{\Theta}_\delta ,$$

$$\boldsymbol{\Sigma}_{yx} = \mathbf{D}_\lambda \boldsymbol{\Phi} ,$$

$$\boldsymbol{\Sigma}_{yy} = \mathbf{D}_\lambda \boldsymbol{\Phi} \mathbf{D}_\lambda + \boldsymbol{\Theta}_\varepsilon ,$$

where $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}_\delta$ and $\boldsymbol{\Theta}_\varepsilon$ are the covariance matrices of $\boldsymbol{\eta}$, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$, respectively. It is seen that $\mathbf{D}_\lambda$ may be multiplied by a nonzero constant, $\boldsymbol{\Phi}$ divided by the same constant and with $\boldsymbol{\Theta}_\delta$ and $\boldsymbol{\Theta}_\varepsilon$ properly adjusted, $\boldsymbol{\Sigma}$ will not change. Hence the model is not identified. One restriction is needed to make it identified but there does not seem to be any meaningful way to choose such a restriction. We shall therefore consider two other models which are both identified (see Jöreskog & Sörbom, 1976a). These models represent different specification of the correlation structures for the errors in $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ as follows:

Model A: The errors are uncorrelated,

Model B: The errors have one common factor.

Model A is shown in Fig. 17 and model B in Fig. 18. In both models

----

Insert Figures 17 and 18 about here

----

the covariance matrix $\underset{\sim}{\Phi}$ of $\underset{\sim}{\eta}$ is restricted to be generated by a simplex or first-order autoregressive model, i.e.

$$\eta_i = \beta_i \eta_{i-1} + \zeta_i \ , \ i = 2,3,4$$

This implies that

$$\underset{\sim}{\Phi} = \begin{bmatrix} \phi_1 & & & \\ \beta_2\phi_1 & \phi_2 & & \\ \beta_2\beta_3\phi_1 & \beta_3\phi_2 & \phi_3 & \\ \beta_2\beta_3\beta_4\phi_1 & \beta_3\beta_4\phi_2 & \beta_4\phi_3 & \phi_4 \end{bmatrix} , \qquad (43)$$

where, as before, $\phi_i = \mathrm{Var}(\eta_i)$, $i = 1,2,3,4$. We now consider the LISREL specification of each of these models. In both models we treat both x and y as dependent variables and use the "no-x" option.

Model A

The LISREL specification is straightforward:

$$\begin{pmatrix} \underset{\sim}{x} \\ \\ \underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{I} \\ \\ \underset{\sim}{D}_\lambda \end{pmatrix} \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \ , \qquad (44)$$

and (40).

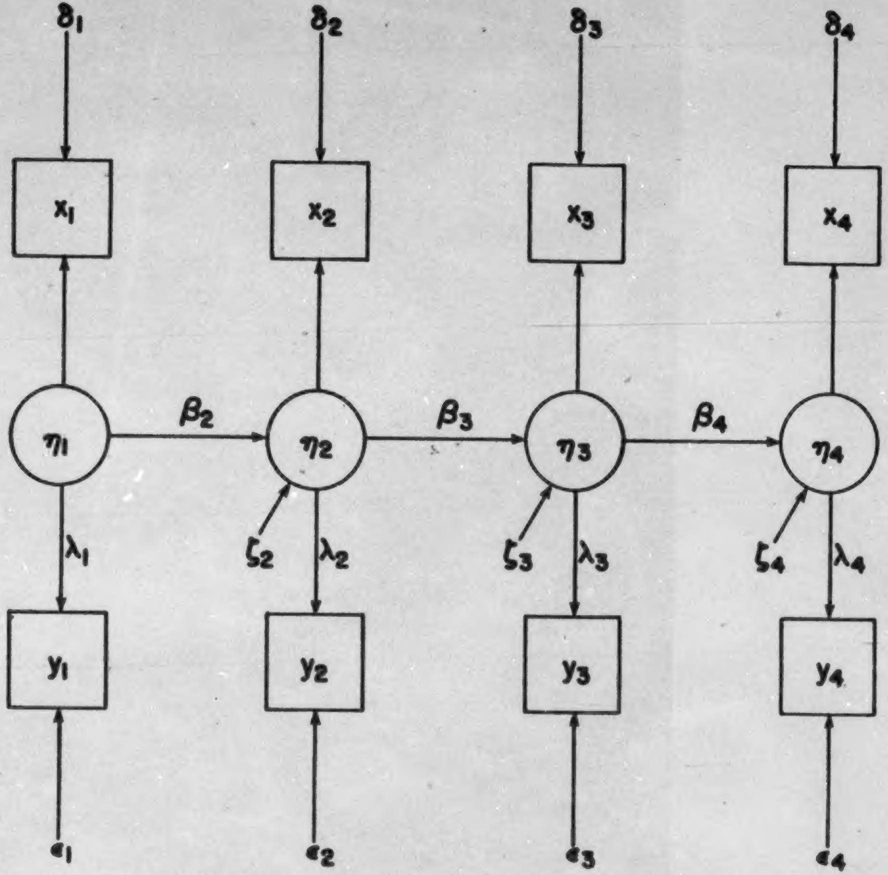

δ1
δ2
δ3
δ4
x1
x2
x3
x4
η1
β2
η2
β3
η3
β4
η4
λ1
ζ2
λ2
ζ3
λ3
ζ4
λ4
y1
y2
y3
y4
ε1
ε2
ε3
ε4

Figure 17 -- A four-wave, two-variable model with uncorrelated errors (Model

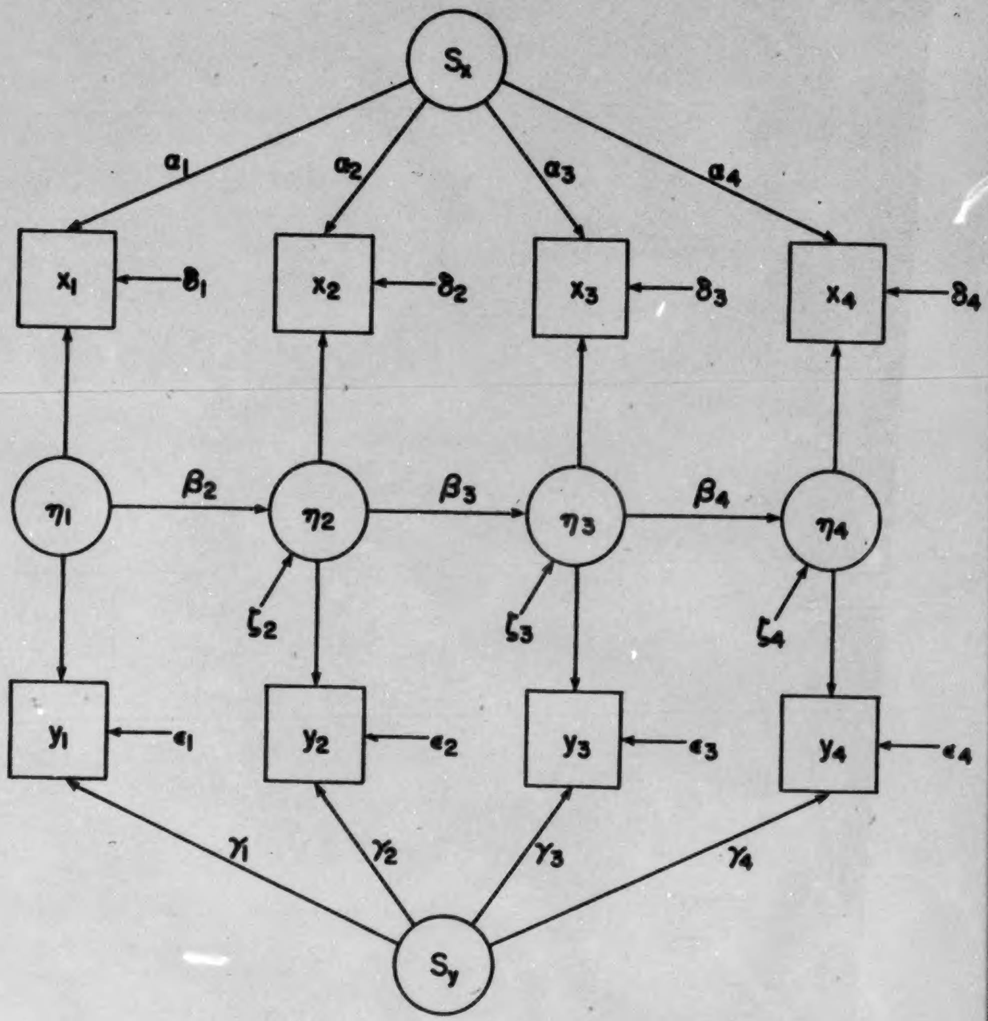

S_x
a_1
a_2
a_3
a_4
x_1
δ_1
x_2
δ_2
x_3
δ_3
x_4
δ_4
η_1
β_2
η_2
β_3
η_3
β_4
η_4
ζ_2
ζ_3
ζ_4
y_1
ϵ_1
y_2
ϵ_2
y_3
ϵ_3
y_4
ϵ_4
γ_1
γ_2
γ_3
γ_4
S_y

Figure 18 -- A four-wave, two-variable model with test-specific factors (Model B)

As before, LISREL treats $\psi_i = \mathrm{Var}(\zeta_i)$ as primary parameters rather than $\phi_i = \mathrm{Var}(\eta_i)$, but estimates of ϕ_i are obtained as a by-product. The one-to-one relationships between ϕ_i and ψ_i, $i=1,2,3,4$, are

$$\phi_1 = \psi_1$$
$$\phi_i = \psi_i + \beta_i^2\phi_{i-1}, \quad i=2,3,4 .$$

The covariance matrix $\underset{\sim}{\Theta}_\varepsilon$ of $\underset{\sim}{\varepsilon}$ is diagonal.

Model B

Model B assumes that the correlations between the errors $\underset{\sim}{\delta}$ and $\underset{\sim}{\varepsilon}$ in Fig. 16 are accounted for by one common factor. These common factors s_x and s_y are test specific factors in contrast to the factors η_1, η_2, η_3 and η_4 which are occasion specific factors in the terminology of Jöreskog (1970a). The test specific factors s_x and s_y are assumed to be uncorrelated and uncorrelated with $\underset{\sim}{\eta}$, $\underset{\sim}{\delta}$ and $\underset{\sim}{\varepsilon}$.

The equations for Model B are

$$\underset{\sim}{x} = \underset{\sim}{\eta} + \underset{\sim}{\alpha}\, s_x + \underset{\sim}{\delta} ,$$

$$\underset{\sim}{y} = D_{\underset{\sim}{\lambda}}\underset{\sim}{\eta} + \underset{\sim}{\gamma}\, s_y + \underset{\sim}{\varepsilon} ,$$

where $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$ are factor loadings relating the observed variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ to the test-specific factors s_x and s_y, respectively. The factors s_x and s_y are scaled to unit variance, for convenience.

Model A is a special case of Model B namely when both $\underset{\sim}{\alpha}$ and $\underset{\sim}{\gamma}$

are zero. The hypothesis $\underset{\sim}{\alpha} = \underset{\sim}{0}$ and $\underset{\sim}{\gamma} = \underset{\sim}{0}$ may be tested with eight degrees of freedom.

The LISREL parameter matrices are specified as

$$\underset{\sim}{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 & 0 & \alpha_1 & 0 \\ 0 & 1 & 0 & 0 & \alpha_2 & 0 \\ 0 & 0 & 1 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & 1 & \alpha_4 & 0 \\ \lambda_1 & 0 & 0 & 0 & 0 & \gamma_1 \\ 0 & \lambda_2 & 0 & 0 & 0 & \gamma_2 \\ 0 & 0 & \lambda_3 & 0 & 0 & \gamma_3 \\ 0 & 0 & 0 & \lambda_4 & 0 & \gamma_4 \end{bmatrix},$$

$$\underset{\sim}{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\underset{\sim}{\Psi} = \mathrm{diag}(\psi_1, \psi_2, \psi_3, \psi_4, 1, 1)$$

where, as before $\psi_i = \mathrm{Var}(\zeta_i)$, i=1,2,3,4 and $\underset{\sim}{\Theta}_\delta$ and $\underset{\sim}{\Theta}_\varepsilon$ are diagonal as before.

C. Estimation of Four-Wave Models for MATH and SCATQ

To illustrate the models of the previous subsection we use the data on the variables MATH and SCATQ from all four occasions. The maximum likelihood estimates of the various parameters are shown in Table 8 along with χ^2-goodness-of-fit-values and corresponding degrees of freedom. It is seen that model A is clearly rejected in favour of model B. The parameters listed in Table 9 are those that come out of the LISREL program and which are used to maximize the likelihood function. Some of these may be very difficult to interpret in a meaningful way. However, from these estimates one can compute various other parmeters which are more easily interpreted. Table 10 gives the estimates of the factor variances and the squared correlations R_i^2 between η_i and η_{i-1} for model B and Table 11 gives the covariance matrices of the errors $\underset{\sim}{\varepsilon}^* = \underset{\sim}{y} - \underset{\sim}{D}_x\underset{\sim}{\eta}$ and $\underset{\sim}{\delta}^* = \underset{\sim}{x} - \underset{\sim}{\eta}$, i.e. the partial covariance matrices of $\underset{\sim}{y}$ and $\underset{\sim}{x}$ after elimination of $\underset{\sim}{\eta}$. Table 12 gives the corresponding correlation matrices.

From Table 10 it is seen that the squared correlations R_i^2 are quite high. There is a very high stability of the quantitative factor over time. This is also indicated by the stability of the β-coefficients in Table 8. Table 11 reveals that covariation among the errors is present for the SCATQ tests to a larger extent than for the MATH tests. Table 12 shows that the correlations among the ε^* s are in general higher than those among the δ^* s. The latter are indeed very small. Hence the model accounts for the intercorrelations among

Insert Tables 9, 10, 11, and 12 about here

TABLE 9

Maximum Likelihood Estimates for Models A and B with x = MATH and y = SCATO

Group: GA (Girls Academic) N = 383

Parameter Estimate	Model A	Model B
$\hat{\lambda}_1$	0.88	0.85
$\hat{\lambda}_2$	1.13	1.06
$\hat{\lambda}_3$	1.23	1.14
$\hat{\lambda}_4$	1.31	1.21
$\hat{\beta}_2$	1.22	1.22
$\hat{\beta}_3$	1.01	1.00
$\hat{\beta}_4$	1.06	1.05
$\hat{\sigma}^2_{\zeta_1}$	55.25	57.31
$\hat{\sigma}^2_{\zeta_2}$	9.28	10.93
$\hat{\sigma}^2_{\zeta_3}$	10.29	13.74
$\hat{\sigma}^2_{\zeta_4}$	2.37	5.79
$\hat{\sigma}_{\delta_1}$	6.19	5.94
$\hat{\sigma}_{\delta_2}$	6.20	5.84
$\hat{\sigma}_{\delta_3}$	5.77	2.29
$\hat{\sigma}_{\delta_4}$	7.55	6.83
$\hat{\sigma}_{\varepsilon_1}$	4.53	4.47
$\hat{\sigma}_{\varepsilon_2}$	6.18	5.83
$\hat{\sigma}_{\varepsilon_3}$	7.40	6.87
$\hat{\sigma}_{\varepsilon_4}$	7.14	6.51
χ^2	72.49	23.18
d.f.	17	9

Additional parameters

Model B

$\hat{\alpha}_1 = 0.97$ $\hat{\gamma}_1 = 1.48$
$\hat{\alpha}_2 = 0.26$ $\hat{\gamma}_2 = 3.51$
$\hat{\alpha}_3 = 4.59$ $\hat{\gamma}_3 = 4.66$
$\hat{\alpha}_4 = -0.92$ $\hat{\gamma}_4 = 4.90$

TABLE 10

Factor Variances and Squared Correlations for Model B

Model B	
$\sigma^2_{\eta i}$	R^2_i
57.31	-
96.23	0.886
109.97	0.875
127.03	0.954

TABLE 11

Covariance Matrices of $\underline{\varepsilon}^*$ and $\underline{\delta}^*$ for Model B

$$\underline{\varepsilon}^* \quad \begin{bmatrix} 22.17 & & & \\ 5.19 & 46.31 & & \\ 6.90 & 16.36 & 68.91 & \\ 7.25 & 17.20 & 22.83 & 66.39 \end{bmatrix}$$

$$\underline{\delta}^* \quad \begin{bmatrix} 36.22 & & & \\ 0.25 & 34.17 & & \\ 4.45 & 1.19 & 26.31 & \\ -0.89 & -0.24 & -4.22 & 47.49 \end{bmatrix}$$

TABLE 12

Correlation Matrices of $\underset{\sim}{\varepsilon}^*$ and $\underset{\sim}{\delta}^*$ for Model B

$$\underset{\sim}{\varepsilon}^* \begin{bmatrix} 1.00 & & & \\ 0.16 & 1.00 & & \\ 0.18 & 0.29 & 1.00 & \\ 0.19 & 0.31 & 0.34 & 1.00 \end{bmatrix}$$

$$\underset{\sim}{\delta}^* \begin{bmatrix} 1.00 & & & \\ 0.01 & 1.00 & & \\ 0.14 & 0.04 & 1.00 & \\ -0.02 & -0.01 & -0.12 & 1.00 \end{bmatrix}$$

the MATH tests much better than the intercorrelations among the SCATQ tests.

D. Multi-wave, two-variables models with background variables

Although the multi-wave, two variables model with freely intercorrelated errors between occasions is not identified, it becomes so as soon as one or more background variables are included. For the case T=2 this was demonstrated in section IV.B. A model with T=4 and two congeneric background variables x_1 and x_2 may be specified as follows. The structural equations are

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \eta_1 \\ -\beta_2 & 1 & 0 & 0 & \eta_2 \\ 0 & -\beta_3 & 1 & 0 & \eta_3 \\ 0 & 0 & -\beta_4 & 1 & \eta_4 \end{bmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix} . \qquad (45)$$

The measurement model for x_1 and x_2 is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_x \end{pmatrix} \xi + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} , \qquad (46)$$

and the measurement model for $\underset{\sim}{y}$ is the same as in (44). The coefficient γ_t measures the direct effect of ξ on η_t and is expected to decrease as t increases. In (46) we have taken ξ to be measured in the same units as x_1. If there is only one background variable x, (46) is replaced by $x = \xi$, i.e., we take $\lambda = 1$ and $\delta = 0$.

E. A General Model for Analysis of Longitudinal Data

In concluding this section we develop a general model for analysis of longitudinal data. All the models considered in the other sections of this chapter are special cases of this general model.

Suppose that several variables are measured at T points in time: $t_1, t_2, \ldots, t_T$, not necessarily equidistant, where time is measured from an arbitrary origin and with an arbitrary unit of measurement. Let p_t dependent variables be measured at occasion t, where t may be $t_1, t_2, \ldots,$ or t_T, and let $y_t' = (y_{1t}, y_{2t}, \ldots, y_{p_t t})$ be a vector of these p_t variables. Neither the number of variables nor the variables themselves need to be the same at all occasions, although in most applications they will be so. At each occasion it is assumed that $\underset{\sim}{y}_t$ has a common factor structure with m_t correlated common factors $\underset{\sim}{\eta}_t' = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{m_t t})$, so that

$$\underset{\sim}{y}_t = \underset{\sim}{\mu}_t + \underset{\sim}{\Lambda}_{yt} \underset{\sim}{\eta}_t + \underset{\sim}{\varepsilon}_t , \tag{47}$$

where $\underset{\sim}{\mu}_t$ is the mean vector of $\underset{\sim}{y}_t$, $\underset{\sim}{\varepsilon}_t$ is a vector of unique factors, and $\underset{\sim}{\Lambda}_{yt}$ is a matrix of order $p_t \times m_t$ of factor loadings.

In addition to the dependent variables $\underset{\sim}{y}_t$, we assume that q independent variables $\underset{\sim}{x}' = (x_1, x_2, \ldots, x_q)$ are measured representing characteristics and conditions existing before the first occasion and assumed to influence the dependent variables $\underset{\sim}{y}_t$. We assume that $\underset{\sim}{x}$ also has a factor structure with common factors $\underset{\sim}{\xi}' = (\xi_1, \xi_2, \ldots, \xi_n)$ so that

$$x = \nu + \Lambda_x + \delta \qquad (48)$$

where ν is the mean vector of x, δ the vector of unique factors and Λ_x the matrix of factor loadings of order qxn.

The structural equations connecting the η's and ξ are assumed to be

$$\eta_1 = A_1 \xi + \zeta_1 \qquad (49)$$

$$\eta_t = A_t \xi + B_t \eta_{t-1} + \zeta_t, \qquad (50)$$

where A_t is a regression matrix of order m_t x n and B_t is a regression matrix of order m_t x m_{t-1}. The vectors $\zeta_t' = (\zeta_{1t}, \zeta_{2t}, \ldots, \zeta_{m_t t})$ are vectors of residuals assumed to be correlated within occasions but uncorrelated between occasions. As before, t may be $t_1, t_2, \ldots, t_T$ and if $t = t_i$ then t-1 is t_{i-1}.

Equations (47) through (50) constitute the basic general model considered in this paper. A special case of this model is when there are no independent variables x. Then equation (48) is no longer included in the model and equations (49) and (50) are replaced by the single equation

$$\eta_t = B_t \eta_{t-1} + \zeta_t. \qquad (51)$$

Equation (47) may be written more compactly as (here illustrated with T = 4 occasions),

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} + \begin{pmatrix} \Lambda_{y1} & 0 & 0 & 0 \\ 0 & \Lambda_{y2} & 0 & 0 \\ 0 & 0 & \Lambda_{y3} & 0 \\ 0 & 0 & 0 & \Lambda_{y4} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \tag{52}$$

and equations (49) and (50) may be combined and written as (in the case of T = 4 occasions)

$$\begin{bmatrix} I & 0 & 0 & 0 \\ -B_2 & I & 0 & 0 \\ 0 & -B_3 & 0 & 0 \\ 0 & 0 & -B_4 & 0 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix} \tag{53}$$

This shows that the model is a LISREL model with

$$y = \mu + \Lambda_y \eta + \varepsilon, \tag{54}$$

$$x = \nu + \Lambda_x \xi + \delta, \tag{55}$$

$$B\eta = \Gamma\xi + \zeta, \tag{56}$$

with $\underset{\sim}{\Lambda}_y$ of order pxm, $\underset{\sim}{B}$ of order mxm and $\underset{\sim}{\Gamma}$ of order mxn, where $p = p_{t_1} + p_{t_2} + \ldots + p_{t_T}$ and $m = m_{t_1} + m_{t_2} + \ldots + m_{t_T}$. In the special case when there is no $\underset{\sim}{x}$, (55) is omitted and (56) should be interpreted as $\underset{\sim}{B}\underset{\sim}{\eta} = \zeta$. Each of the matrices $\underset{\sim}{\Lambda}_x$, $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$ and $\underset{\sim}{\Lambda}_{yt}$, t = 1, 2, ..., T may contain fixed, free and constrained parameters as in section II.D.

VI. SUMMARY

In this chapter we have developed several models suitable for analyzing longitudinal data and considered the statistical problems of model specification, identification, estimation and testing. Almost all of the models may be estimated and tested using three computer programs, ACOVS, LISREL and COFAMM, which are described briefly in section II.

Section III deals with the estimation of polynomial growth curves describing the means of response variables as functions of time. The growth curves may be estimated for several variables and for several groups of individuals simultaneously and various hypotheses may be tested such that (i) the growth curve has a specified degree, (ii) the growth curves are identical or parallel for several variables and/or groups. The estimation of growth curves when the response variables are auto-regressive is also considered. If the auto-regressive model holds, the growth curves can be estimated more efficiently and the tests will be more powerful.

Sections IV and V deal with models involving latent variables or hypothetical constructs and the related problem of measurement errors in the observed variables. The kernel in these models is a set of linear structural relationships among latent variables that are not directly observed but observed by means of two or more indicators. We consider models with or without background variables. Section IV deals with two-wave models and section V with multi-wave models. In subsections IV.E-F, we consider the comparison of change between groups under the assumption that certain characteristics of the response variables are invariant over groups of people.

For most of the models in sections IV and V we consider the identification problem. The estimation problem is considered in the sense that it is shown how to specify the model for one of the three computer programs described in section II. For some of the models, the estimation and testing is illustrated by some data.

REFERENCES

Ambron, S. R., & Rogosa, D. R. Structural equation models and causal inference in child development research. Paper presented at the American Educational Research Association Convention, Washington, D. C., April 1975

Andersen, E. B., & Madsen, M. Estimating the parameters of the latent population distribution. *Institute of Statistics Research Report No. 35*, University of Copenhagen, 1976.

Anderson, R. L., & Bancroft, T. A. *Statistical theory in research*. New York: McGraw-Hill, 1952.

Anderson, S. B., & Maier, M. H. 34,000 pupils and how they grew. *Journal of Teacher Education*, 1963, *14*, 212-216.

Anderson, T. W. *An introduction to multivariate statistical analysis*. New York: Wiley, 1958.

Anderson, T. W. Some stochastic process models for intelligence test scores. In K. J. Arrow, S. Karlin & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959*. Stanford, Calif.: Stanford University Press, 1960.

Ausubel, D., & Sullivan, E. *Theories and problems of child development*. New York: Grune and Stratton, 1970.

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145-171.

Baltes, P. B. Prototypical paradigms and questions in life-span research on development and aging. *Gerontologist*, 1973, *13*, 458-467.

Baltes, P. B. & Nesselroade, J. R. Culture change and adolescent personality development. *Developmental Psychology*, 1972, *7*, 244-256.

Baltes, P. B., & Nesselroade, J. R. The developmental analysis of individual differences on multiple measures. In. J. R. Nesselroade & H. W. Reese (Eds.) *Life-span developmental psychology: Methodological Issues*. New York: Academic Press, 1973.

Baltes, P. B., & Nesselroade, J. R. A developmentalist's view of regression toward the mean: A largely irrelevant issue in the study of developmental change? Unpublished manuscript, The Pennsylvania State University, 1976.

Baltes, P. B., Nesselroade, J. R., Schaie, K. W., & Labouvie, E. W. On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology*, 1972, *6*, 78-84.

Baltes, P. B., & Willis, S. L. Toward psychological theories of aging and development. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand-Reinhold, 1976.

Bargmann, R. E. A study of independence and dependence in multivariate normal analysis. *Mimeo Series No. 186*, Institute of Statistics, University of North Carolina at Chapel Hill, 1957.

Bentler, P. M. Assessment of developmental factor change at the individual and group level. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues*. New York: Academic Press, 1973.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Eds.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Bergman, L. R. Some univariate models in studying change. *Reports from the Psychological Laboratories*, Supplement 10, 1971, University of Stockholm.

Berkson, J. Application of the logistic function to bioassay. *Journal of the American Statistical Association*, 1944, *39*, 357-365.

Berkson, J. A statistically precise and relatively simple method of estimating the bioassay with quantal response based on the logistic function. *Journal of the American Statistical Association*, 1953, *48*, 565-599.

Berkson, J. Maximum likelihood and minimum χ^2 estimates of the logistic function. *Journal of the American Statistical Association*, 1955, *50*, 130-162.

Bhapkar, V. P. A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, 1966, *61*, 228-235.

Bhapkar, V. P. On the analysis of contingency tables with a quantitative response. *Biometrics*, 1968, *24*, 329-338.

Bhapkar, V. P., & Koch, G. G. Hypotheses of "no interaction" in multidimensional contingency tables. *Technometrics*, 1968, *10*, 107-123. (a)

Bhapkar, V. P., & Koch G. G. On the hypotheses of "no interaction" in multidimensional contingency tables. *Biometrics*, 1968, *24*, 567-594. (b)

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. *Discrete multivariate analysis: Theory and practice*. Cambridge, Massachusetts: M.I.T. Press, 1975.

Blalock, H. M. Making causal inferences for unmeasured variables from correlations among indicators. *American Journal of Sociology*, 1963, *69*, 53-62.

Blalock, H. M. (Ed.) *Measurement in the social sciences*. Chicago: Aldine, 1974.

Bock, R. D. Multivariate analysis of variance of repeated measurements. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 1972, *37*, 29-51.

Bock, R. D. Basic issues in the measurement of change. In D. N. M. deGruijter and L. J. T. van der Kamp (Eds.), *Advances in psychological and educational measurement*. London: Wiley & Sons, 1976.

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.

Bock, R. D., & Repp, B. H. *MATCAL: Double-precision matrix operations subroutines for the IBM System/360-370 computers.* Chicago: National Educational Resources, 1974.

Bock, R. D., & Thissen, D. Fitting multi-component models for growth in stature. *Proceedings of the 9th International Biometric Conference*, 1977, in press.

Bock, R. D., & Thrash, W. Characterizing a latent trait distribution. *Proceedings of the 1976 Dayton Symposium on Application of Statistics*, 1977, in press.

Bock, R. D., Wainer, H., Petersen, A., Murray, J., & Roche, A. A parameterization for individual human growth curves. *Human Biology*, 1973, *45*, 63-80.

Bohrnstedt, G. W. Observations on the measurement of change. In E. F. Borgatta (Ed.). *Sociological methodology: 1969*. San Francisco: Jossey-Bass, 1969.

Bolz, C. R. Types of personality. In R. M. Dreger (Ed.), *Multivariate personality research: Contributions to the understanding of personality in honor of Raymond B. Cattell*. Baton Rouge: Claitor, 1972.

Bowes, S., & Gintis, H. I.Q. in the U. S. class structure. *Social Policy*, 1972, *3*, 63-96.

Bowker, A. H. A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 1948, *43*, 572-574.

Box, G. E. P. Problems in the analysis of growth and wear curves. *Biometrics*, 1950, *6*, 362-389.

Box, G. E. P. Bayesian approaches to some bothersome problems in data analysis. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis*. Chicago: Rand McNally, 1967.

Bracht, G. H., & Glass, G. V. The external validity of experiments. *American Eudcational Research Journal*, 1968, *5*, 437-474.

Bronfenbrenner, U. Developmental research, public policy, and the ecology of childhood. *Child Development*, 1974, *45*, 1-5.

Bronfenbrenner, U. The experimental ecology of education. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1976.

Browne, E. T. *Introduction to the theory of determinants and matrices*. Chapel Hill, N.C.: University of North Carolina Press, 1958.

Buss, A. R. A general developmental model for interindividual differences, intraindividual differences and intraindividual changes. *Developmental Psychology*, 1974, *10*, 70-78. (a)

Buss, A. R. Multivariate model of quantitative, structural, and quantistructural ontogenetic change. *Developmental Psychology*, 1974, *10*, 190-203. (b)

Cahan, L. S., & Linn, R. L. Regions of significant criterion differences in aptitude-treatment-interaction research. *American Educational Research Journal*, 1971, *8*, 521-530.

Campbell, D. T. From description to experimentation: Interpreting trends as quasiexperiments. In C. W. Harris (Ed.) *Problems in Measuring Change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Campbell, D. T., & Erlebacher, A. How regression artifacts in quasiexperimental evaluations can mistakenly make compensatory education look harmful. In Hellmuth, J. (Ed.) *Compensatory education: A national debate, Vol. III. The disadvantaged child*. New York: Brunner-Mazel, 1970.

Cattell, R. B. *Description and measurement of personality*. Yonkers, N.Y.: World Book, 1946.

Cattell, R. B. The three basic factor-analytic research designs: Their interrelations and derivatives. *Psychological Bulletin*, 1952, *49*, 499-520.

Cattell, R. B. The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966.

Cattell, R. B., Coulter, M. A., & Tsujioka, B. The taxonometric recognition of types and functional emergents. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966.

Cicchetti, D. V. A new measure of agreement between rank-ordered variables. *Proceedings of the 80th Annual Convention of the American Psychological Association*, 1972, *7*, 17-18. (Summary)

Clark, D. E., Clarke, A. M., & Brown, R. I. Regression to the mean--a confused concept. British Journal of Psychology, 1960, 51, 105-117.

Clarke-Stewart, K. A. Interactions between mothers and their young children: Characteristics and consequences. Monographs of the Society for Research in Child Development, 1973, 38 (Serial No. 153).

Clarke-Stewart, K. A. Dealing with the complexity of mother-child interaction. Paper presented at the biennial meeting of the Society for Research in Child Development, Denver, April 1975.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 1968, 70, 213-220.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N. J.: Erlbaum Associates, 1975.

Cole, J. W. L., & Grizzle, J. E. Application of multivariate analysis of variance to repeated measurements experiments. Biometrics, 1966, 22, 810-827.

Coleman, J. S. Introduction to mathematical sociology. New York: Free Press, 1964.

Coleman, J. S. The mathematical study of change. In H. M. Blalock (Ed.), Methodology in social research. New York: McGraw-Hill, 1968.

Coleman, J. S., Katz, E. & Menzel. H. The diffusion of an innovation among physicians. Sociometry, 1957, 20, 253-270.

Corballis, M. C. A factor model for analysing change. The British Journal of Mathematical & Statistical Psychology, 1973, 26, 90-97.

Corballis, M. C., & Traub, R. E. Longitudinal factor analysis. Psychometrika, 1970, 35, 79-98.

Count, E. W. Growth patterns of human physique. Human Biology, 1943, 15, 1-32.

Crano, W. D., Kenny, D. A., & Campbell, D. T. Does intelligence cause achievement?: A cross-lagged panel analysis. Journal of Educational Psychology, 1972, 63, 258-275.

Cronbach, L. J., & Furby, L. How should we measure "change" - or should we? Psychological Bulletin, 1970, 75, 68-80.

Damon, A. Discrepancies between findings of longitudinal and cross-sectional studies in adult life: Physique and physiology. Human Development, 1965, 8, 16-22.

Dayton, C. M. Design of educational experiments. New York: McGraw-hill, 1970.

Deakin, M. A. B. Compertz curves, allometry and embryogenesis. Bulletin Mathematical Biophysics, 1970, 32, 445-452.

DeLury, D. B. Values and integrals of the orthogonal polynomials up to n=26. Toronto: University of Toronto Press, 1950.

Deming, J. Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth. Human Biology, 1957, 29, 83-122.

Duncan, O. D. Path analysis: Sociological examples. American Journal of Sociology, 1966, 72, 1-16.

Duncan, O. D. Contingencies in constructing causal models. In E. Borgatta (Ed.), Sociological methodology 1969. San Francisco: Jossey-Bass, 1969.

Duncan, O. D. Some linear models for two wave, two-variable panel analysis. Psychological Bulletin, 1969, 72, 177-182.

Duncan, O. D. Unmeasured variables in linear models for panel analysis. In H. L. Costner (Ed.), Sociological Methodology: 1972, San Francisco: Jossey-Bass, 1972.

Duncan, O. D. Introduction to structural equation models. New York: Academic Press, 1975. (a)

Duncan, O. D. Some linear models for two-wave, two-variable panel analysis with one-way causation and measurement errors. In H. M. Blalock Ed.), Quantitative sociology: International perspectives on mathematical and statistical modelling. New York: Academic Press, Inc., 1975. (b)

Einhorn, H. J. Alchemy in the behavioral sciences. The Public Opinion Quarterly, 1972, 36, 367-378.

Emmerich, W. Personality development and concepts of structure. Child Development, 1968, 39, 671-690.

Evans, S. H., & Anastasio, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. Psychological Bulletin, 1968, 69, 225-234.

Federer, W. T. Experimental design; Theory and application. New York: Macmillan, 1955.

Feller, W. Introduction to probability theory and its applications (Vol. 1). New York: Wiley, 1968.

Feller, W. An introduction to probability and it applications (Vol. 2). New York: Wiley, 1966.

Feller, W. *Introduction to probability theory and its applications* (Vol. 2). New York: Wiley, 1971.

Ferguson, G. A. *Statistical analysis in psychology and education* (2nd Ed.) New York: McGraw-Hill, 1966.

Finn, J. D. *Multivariance: Univariate and multivariate analysis of variance, covariance, and regression*. Chicago: National Educational Resources, 1974.

Finn, J. D. Personal communication, 1976.

Finn, J. D., & Mattsson, I. *Multivariate analysis in educational research: Applications of the MULTIVARIANCE program*. Chicago: National Educational Resources, 1977.

Finney, D. J. *Statistical method in biological assay* (2nd Edition). London: Charles Griffin, 1964.

Fisher, R. A. Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *Journal of Agricultural Science*, 1921, *11*, 107-135.

Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research* (6th ed.). New York: Hafner, 1963.

Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 1971, *76*, 378-382.

Forthofer, R. N., & Koch, G. G. An analysis for compounded functions of categorical data. *Biometrics*, 1973, *29*, 143-157.

Frederiksen, C. R. Models for the analysis of alternative sources of growth in correlated stochastic variables. *Psychometrika*, 1974, *39*, 223-245.

Freeman, J., & Hannan, M. T. Growth and decline processes in organizations. *American Sociological Review*, 1975, *40*, 215-228.

Furby, L. Interpreting regression toward the mean in developmental research. *Developmental Psychology*, 1973, *8*, 172-179.

Gaito, J., & Wiley, D. E. Univariate analysis of variance procedures in measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Games, P. A. Limitations of analysis of covariance on intact group quasi-experimental designs. *Journal of Experimental Education*, 1976, *44*(4), 51-53.

Games, P. A. & Klare, G. R. *Elementary statistics: Data analysis for the behavioral sciences*. New York: McGraw-Hill, 1967.

Gantmacher, F. R. *Theory of matrices* (Vol. 1). New York: Chelsea, 1960.

Geisser, S. A Bayes approach for combining correlated estimates. *Journal of the American Statistical Association*, 1965, *60*, 602-607.

Geisser, S., & Kappenman, R. F. A posterior region for parallel profile differentials. *Psychometrika*, 1971, *36*, 71-78.

Gesell, A. The ontogenesis of infant behavior. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954.

Ghosh, M., Grizzle, J. E. & Sen, P. K. Nonparametric methods in longitudinal studies. *Journal of the American Statistical Association*, 1973, *68*, 29-36.

Ginsberg, R. Semi-Markov processes and mobility, *Journal of Mathematical Sociology*, 1971, *1*, 233-263.

Goldberger, A. S. Econometrics and psychometrics: A survey of communalities. Psychometrika, 1971, *36*, 83-108.

Goldberger, A. S. Structural equation methods in the social sciences. *Econometrica*, 1972, *40*, 979-1001

Goldberger, A. Structural equation models: An overview. In A. Goldberger & O. D. Duncan (Eds.) *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Gompertz, B. On the nature of the function expressive of the law of human mentality. *Philosophical Transactions of the Royal Society*, 1825, *115*, 513-583.

Goodman, L. A. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, 1974, *79*, 1179-1259.

Gordon, R. A. Issues in multiple regression. *American Journal of Sociology*, 1968, *73*, 592-616.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometrika*, 1959, *24*, 95-112.

Grizzle, J. E., & Allen D. M. Analysis of growth and dose response curves. Biometrics, 1969, *25*, 307-318.

Grizzle, J. E., Starmer, C. F., & Koch, G. G. Analysis of categorical data by linear models. *Biometrics*, 1969, *25*, 489-504.

Grizzle, J. E., & Williams, O. D. Log linear models and tests of independence for contingency tables. *Biometrics*, 1972, *28*, 137-156.

Guttman, L. A. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Columbia University Press, 1954.

Hald, A. *Statistical tables and formulas*. New York: Wiley, 1952.

Hannan, M. T., Rubinson, R., & Warren, J. T. The causal approach to measurement error in panel analysis: Some further contingencies. In H. M. Blalock (Ed.), *Measurement in the social science*. Chicago: Aldine, 1974.

Hannah, M. T., & Young, A. A. Estimation in panel models: Results on pooling cross-sections and time-series. Technical Report #51, Stanford University, Laboratory for Social Research, 1974.

Härnqvist, K. Relative changes in intelligence from 13 to 18. *Scandinavian Journal of Psychology*, 1968, *9*, 50-82.

Harris, C. W. Canonical factor models for the description of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Harris, C. W. (Ed.) *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Heise, D. R. Separating reliability and stability in test-retest correlations. *American Sociological Review*, 1969, *34*, 93-101.

Heise, D. R. Causal inference from panel data. In E. F. Borgatta and G. W. Bohrnstedt (Eds.), *Sociological methodology: 1970*. San Francisco: Jossey-Bass, 1970.

Heise, D. R. *Causal analysis*. New York: Wiley, 1975.

Heller, C. A., Scott, E. M., & Hammes, L. M. Height, weight and growth of Alaskan Eskimos. *American Journal of Diseases of Childhood*, 1967, *113*, 338-344.

Hilton, T. L. *Growth study annotated bibliography: Progress report 69-11*. Princeton, N.J.: Educational Testing Service, 1969.

Hindley, C. B. The place of longitudinal methods in the study of development. In F. J. Monks, W. W. Hartup, and J. de Wit (Eds.), *Determinants of behavioral development*. New York: Academic Press, 1972.

Hoem, J. Inhomogeneous semi-Markov processes, select acturial tables, and duration dependence in demography. In T. N. E. Greville (Ed.), *Population dynamics*. New York: Academic Press, 1972.

Horn, J. L. Significance tests for use with r_p and related profile statistics. *Educational and psychological development*, 1961, *21*, 363-370.

Huston-Stein, A., & Baltes, P. B. Theory and method in life-span developmental psychology: Implications for child development. In H. W. Reese & L. P. Lipsitt (Eds.), Advances in child development and behavior. New York: Academic Press, in press.

Huynh, H. & Feldt, L. S. Conditions under which mean square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582-1589.

Israelsohn, W. J. Description and modes of analysis of human growth. In J. M. Tanner (Ed.), *Human growth*. New York: Pergamon Press, 1960.

Jamison, P. L. Anthropometric variation. In P. L. Jamison & S. L. Zegura (Eds.), *The Eskimos of north western Alaska: A biological perspective*. Stroudsburg, Pa.: Dowden, Hutchison and Ross, in press.

Jenss, R. M. & Bayley, N. A mathematical method for studying growth in children. *Human Biology*, 1937, *9*, 556-563.

Jones, H. E. Problems of method in longitudinal research. *Vita Humana*, 1958, *1*, 93-99.

Jöreskog, K. G. Factoring the multitest-multioccasion correlation matrix. In C. E. Lunneborg (Ed.), *Current problems and techniques in multivariate psychology*. (Proceedings of a Conference Honoring Professor Paul Horst.) Seattle: University of Washington, 1970. (a)

Jöreskog, K. G. Estimation and testing of simplex models. *The British Journal of Mathematical and Statistical Psychology*, 1970, *23*, 121-145. (b)

Jöreskog, K. G. A general method for analysis of covariance structures. *Biometrika*, 1970, *57*, 239-251. (c)

Jöreskog, K. G. Simultaneous factor analysis in several populations. *Psychometrika*, 1971, *36*, 409-426.

Jöreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press, 1973. (a)

Jöreskog, K. G. Analysis of covariance structures. In P. R. Krishnaiah (Ed.), *Multivariate analysis - III*. New York: Academic Press, Inc., 1973. (b)

Jöreskog, K. G. Structural equation models in the social sciences: Specification, estimation and testing. Research Report 76-9. University of Uppsala, Department of Statistics, 1976.

Jöreskog, K. G. & Sörbom, D. Some models and estimation methods for analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models*. Amsterdam: North Holland Publishing Co., in press, 1976. (a)

Jöreskog, K. G. & Sörbom, D. Statistical models and methods for analysis of longitudinal data. Research Report 76-1, Department of Statistics, University of Uppsala, Sweden, 1976.

Jöreskog, K. G. & Sörbom, D. Statistical models and methods for test-retest situations. In D.N.M. de Gruijter, L. J. Th. van der Kamp & H. F. Crombag (Eds.), *Advances in psychological and educational measurement*. London: Wiley, 1976. (b)

Jöreskog, K. G. & Sörbom. D. *LISREL - Estimation of linear structural equation systems by maximum likelihood methods*. Chicago: International Eudcational Services, 1976. (c)

Jöreskog, K. G., van Thillo, M., & Gruvaeus, G. T. ACOVSM-A general computer program for analysis of covariance structures including generalized MANOVA. Res. Bull. 70-01. Educational Testing Service: Princeton, New Jersey, 1971.

Kaplan, B. The study of language in psychiatry. In S. Arieti (Ed.), *American handbook of psychiatry* (Vol. 3). New York: Basic Books, 1966.

Kendall, M. G., & Stuart, A. *The Advanced theory of statistics*, Vol. 2, *Inference and relationship*. Griffin: London, 1961.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics*. (Vol. 3, 2nd Ed.) London: Griffin, 1968.

Kenny, D. A. Cross-lagged and synchrononous common factors in panel data. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Kessen, W. Research design in the study of developmental problems. In P. H. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960.

Kessen. W. Stage and structure in the study of children. *Monographs of the Society for Research in Child Development*, 1962, *27*, 65-81.

Khatri, C. G. A note on a MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics*, 1966, 18, 75-86.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, California: Brooks/Cole, 1968.

Kleinbaum, D. G. A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, 1973, *3*, 117-124.

Koch, G. G. Some aspects of the statistical analysis of 'split plot' experiments in completely randomized layouts. *Journal of the American Statistical Association*, 1969, *64*, 485-505.

Koch, G. G. The use of non-parametric methods in the statistical analysis of a complex split plot experiment. Biometrics, 1970, *26*, 105-128.

Koch, G. G., Freeman, J. L., Freeman, D. H., Jr., Landis, J. R. & Lehnen, R. G. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 1977, in press.

Koch, G. G., Imrey, P. B., Freeman, D. H., Jr., & Tolley, H. D. The asymptotic covariance structure of estimated parameters from contingency table log-linear models. Proceedings of the 9th International Biometric Conference, Boston, August 1976.

Koch, G. G., Imrey, P. B., & Reinfurt, D. W. Linear model analysis of categorical data with incomplete response vectors. Biometrics, 1972, 28, 663-692.

Koch, G. G., Johnson, W. D., & Tolley, H. D. A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. Journal of the American Statistical Association, 1972, 67, 783-796.

Koch, G. G., & Reinfurt, D. W. The analysis of categorical data from mixed models. Biometrics, 1971, 27, 157-173.

Kohlberg, L. Early education: A cognitive-developmental view. Child Development, 1968, 39, 1013-1061.

Kohlberg, L. Continuities in childhood and adult moral development revisited. In P. B. Baltes & K. W. Schaie (Eds.), Life-span developmental psychology: Personality and socialization, New York: Academic Press, 1973.

Kowalski, C. J. The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. Technometrics, 1970, 12, 517-544.

Kowalski, C. J. A commentary on the use of multivariate statistical methods in anthropometric research. American Journal of Physical Anthropology, 1972, 36, 119-131.

Kowalski, C. J., & Guire, K. E. Longitudinal data analysis. Growth, 1974, 38, 131-169.

Kuhlen, R. G. Age and intelligence: The significance of cultural change in longitudinal vs. cross-sectional findings. Vita Humana, 1963, 6, 113-124.

Labouvie, E. W. Developmental causal structures of organism-environment interactions. Human Development, 1974, 17, 444-452.

Labouvie, E. W., Bartsch, T. W., Nesselroade, J. R., & Baltes, P. B. On the internal and external validity of simple longitudinal designs. Child Development, 1974, 45, 282-290.

Landis, J. R. A general methodology for the measurement of observer agreement when the data are categorical. Unpublished doctoral dissertation, University of North Carolina, 1975.

Landis, J. R. & Koch, G. G. A review of statistical methods in the analysis of data arising from observer reliability studies (Part I). Statistica Neerlandica, 1975, 29, 101-123. (a)

Landis, J. R. & Koch, G. G. A review of statistical methods in the analysis of data arising from observer reliability studies (Part II). *Statistica Neerlandica*, 1975, *29*, 151-161. (b)

Landis, J. R., & Koch, G. G. The measurement of majority agreement among many observers in a clinical diagnosis example. *Biometrics*, 1977, in press. (a)

Landis, J. R., & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, 1977, in press. (b)

Landis, J. R., Stanish, W. M., & Koch, G. G. A computer program for the generalized chi-square analysis of categorical data using weighted least squares (GENCAT). Unpublished manuscript, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, 1976. (To appear in *Computer Programs in Biomedicine*.)

Lee, W. *Experimental design and analysis*. San Francisco: Freeman & Co., 1975.

Lee, Y. K. A note on Rao's reduction of Pottoff and Roy's generalized linear model. *Biometrika*, 1974, *61*, 349-351.

Lewis, M., & Rosenblum, L. (Eds.) *The effect of the infant on its caregivers*. New York: Wiley, 1974.

Light, R. J. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, 1971, *76*, 365-377.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.

Loevinger, J. The meaning and measurement of ego development. *American Psychologist*, 1966, *21*, 195-206.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wisconsin: University of Wisconsin Press, 1963.

Lord, F. M., & Novick. M. N. *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley Publishing Company, 1968.

Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 1974, *39*, 247-264.

Madaus, G. F., Woods, E. M., & Nuttall, R. L. A causal model analysis of Bloom's taxonomy. *American Educational Research Journal*, 1973, *10*, 253-262.

McCall, R. B., & Appelbaum, M. I. Bias in the analysis of repeated measures designs: Some alternative approaches. *Child Development*, 1973, *44*, 401-415.

McFarland, D. D. Intra-generational social mobility as a Markov process: Including a time-stationary Markovian model that explains declines in mobility rates over time. American Sociological Review, 1970, 35, 463-476.

McGinnis, R. A stochastic model of social mobility. American Sociological Review, 1968, 33, 712-722.

McKeon, J. J. F approximations to Hotelling's T_0^2. Biometrika, 1974, 61, 381-383.

McNemar, Q. Psychological Statistics (4th Ed.) New York: Wiley, 1969.

Morrison, D. F. Multivariate statistical methods. New York: McGraw Hill, 1967.

Morrison, D. F. The analysis of a single sample of repeated measurements. Biometrics, 1972, 28, 55-71.

Myers, J. L. Fundamentals of experimental design. (2nd ed.) Boston: Allyn & Bacon, 1972.

Nesselroade, J. R. Application of multivariate strategies to problems of measuring and structuring long-term change. In L. R. Goulet & P. B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970.

Nesselroade, J. R. Note on the "Longitudinal factor analysis" model. Psychometrika, 1972, 37, 187-191.

Nesselroade, J. R., & Reese, H. W. Life-span developmental psychology: Methodological issues. New York: Academic Press, 1973.

Neyman, J. Contribution to the theory of the χ^2 test. In J. Neyman (Ed.), Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. Berkeley and Los Angeles: University of California Press, 1949.

Olsson, U., & Bergman, L. A. A longitudinal factor model for studying change in ability structure. Multivariate Behavioral Research, 1977, in press.

O'Mally, P. M., & Bachman, J. G. Self-esteem and educational attainment: A longitudinal analysis. Paper presented at the Annual Meering of the American Educational Research Association, San Francisco, April 1976.

Pearson, E. S., & Hartley, H. O. Biometrika tables for statisticians (Vol. 1, 3rd ed). Cambridge University Press, 1966.

Phillips, D. C. & Kelley, M. E. Hierarchical theories of development in education and psychology. Harvard Educational Review, 1975, 45, 351-375.

Piaget, J. The construction of reality in the child. New York: Basic Books, 1954.

Piaget, J. The general problem of the psychobiological development of the child. In J. M. Tanner & B. Inhelder (Eds.), *Discussions on child development* (Vol. 4). New York: International University Press, 1960.

Pillai, K. C. S. *Statistical tables for tests of multivariate hypotheses*. Manila: Statistical Center, University of the Philippines, 1960.

Pillai, K. C. S., & Jayachandran, K. On the exact distribution of Pillai's $V^{(s)}$ criterion. *Journal of the American Statistical Association*, 1970, 65, 447-454.

Popper, K. R. *Conjectures and refutations* (4th ed.). London: Routledge & Kegan, 1972.

Pottoff, R. F., & Roy, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 1964, *51*, 313-326.

Prahl-Anderson, B., & Kowalski, C. J. A mixed longitudinal interdisciplinary study of the growth and development of Dutch children. *Growth*, 1973, *37*, 281-295.

Purkey, W. W. *Self concept and school achievement*. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Radin, N. Maternal warmth, achievement motivation, and cognitive functioning in lower-class preschool children. *Child Development*, 1971, *42*, 1560-1565.

Radin, N. Observed maternal behavior with four-year-old boys and girls in lower-class families. *Child Development*, 1974, *45*, 1126-1131.

Rao, C. R. Some statistical methods for the comparison of growth curves. *Biometrics*, 1958, *14*, 1-17.

Rao, C. R. Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, 1959, *46*, 49-58.

Rao, C. R. The theory of least squares when parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 1965, *52*, 447-458.

Rao, C. R. Covariance adjustment and related problems in multivariate analysis. In P R. Krishnaiah (Ed.), *Multivariate analysis*. New York: Academic Press, 1966.

Rao, C. R. Least squares theory using an estimated dispersion matirx and its applications to measurement of signals. *Proceedings of the 5th Berkeley Symposium*, 1967, *1*, 355-372.

Rao, M. N., & Rao, C. R. Linked cross sectional study for determining norms and growth rates--a pilot survey of Indian school-going boys. *Sankhya B*, 1966, *28*, 237-258.

Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Institute of Mathematics and Statistics, University of Copenhagen, 1960.

Reese, H. W. The scope of experimental child psychology. In H. W. Reese & L. P. Lipsitt (Eds.), *Experimental child psychology*. New York: Academic Press, 1970.

Rogosa, D. R. & Ambron, S. R. Analyzing reciprocal effects in child development: Applications to attachment. Paper presented at the Western Psychological Association Convention, Los Angeles, April 1976.

Rogosa, D. R., Webb, N., & Radin, N. Causal models in child development: A reanalysis of data on cognitive development in lower-class children. Unpublished manuscript, Department of Education, Stanford University, 1976.

Roy, J. Step-down procedure in multivariate analysis. *Annals of Mathematical Statistics*, 1958, *29*, 1177-1187.

Roy, S. N. *Some aspects of multivariate analysis*. New York: Wiley, 1957.

Roy, S. N., & Bargmann, R. E. Tests of multiple independence and the associated confidence bounds. *Annals of Mathematical Statistics*, 1958, *29*, 491-503.

Rubin, D. B. Characterizing the estimation of parameters in incomplete data problems. *Journal of the American Statistical Association*, 1974, *69*, 467-474.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17*, 1969.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92-107.

Schaie, K. W. A reinterpretation of age-related changes in cognitive structure and functioning. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory*. New York: Academic Press, 1970.

Schaie, K. W. Can the longitudinal method be applied to the study of psychological development? In F. J. Monks, W. W. Hartup, & J. de Wit (Eds.), *Determinants of behavioral development*. New York: Academic Press, 1972.

Schock, N. W. Growth curves. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Shirley, M. M. *The first two years: Personality manifestations* (Vol. 1). Minneapolis: University of Minnesota, 1933.

Silvey, S. D. *Statistical Inference*. Middlesex: Penguin Books, 1970.

Simon, H. Spurious correlations: A causal interpretation. Journal of the American Statistical Association, 1954, 49, 467-479.

Singer, B., & Spilerman, S. Social mobility models for heterogeneous populations. In H. Costner (Ed.), Sociological methodology 1973-74. San Francisco: Jossey-Bass, 1974.

Singer, B., & Spilerman, S. The representation of social processes by Markov models. American Journal of Sociology, 1976, 82, 1-54.

Sörbom, D. A general method for studying differences in factor means and factor structure between groups. The British Journal of Mathematical & Statistical Psychology, 1974, 27, 229-239.

Sörbom, D. Detection of correlated errors in longitudinal data. The British Journal of Mathematical & Statistical Psychology, 1975, 28, 138-151.

Sörbom, D. & Jöreskog, K. G. Confirmatory factor analysis with model modification. Chicago: International Educational Services, 1976.

Spilerman, S. Extensions of the mover-stayer model. American Journal of Sociology, 1972, 78, 599-626. (a)

Spilerman, S. The analysis of mobility processes by the introduction of independent variables into a Markov chain. American Sociological Review, 1972, 37, 277-294. (b)

Steel, R. G. D., & Torrie, J. H. Principles and procedures of statistics. New York: McGraw-Hill, 1960.

Summers, G. F., Hough, R. L., Scott, J. T. & Folse, C. L. Before industrialization: A rural social system base study. Bulletin No. 736. Urbana, Illinois: Illinois Agricultural Experiment Station, University of Illinois, 1969.

Suppes, P. Facts and fantasies of education. In M. C. Wittrock (Ed.), Changing education: Alternatives from educational research. Englewood Cliffs, New Jersey: Prentice Hall, 1973.

Suppes, P. The place of theory in educational research. Educational Researcher, 1974, 3, 3-10.

Thissen, D. M. Incorporating item response latencies in latent trait estimation. Unpublished Ph.D. dissertation, Department of Behavioral Sciences, University of Chicago, 1976.

Thissen, D., Bock, R. D., Wainer, H., & Roche, A. F. Individual growth in stature: A comparison of four growth studies in the U.S.A. Annals of Human Biology, 1976, 3, 529-542.

Thorndike, R. L. Intellectual status and intellectual growth. Journal of Educational Psychology, 1966, 57, 121-127.

Timm, N. H. *Multivariate analysis with applications in education and psychology*. Monterey, Calif.: Brooks/Cole, 1975.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Trawinski, I. M., & Bargmann, R. E. Maximum likelihood estimation with incomplete multivariate data. *Annals of Mathematical Statistics*, 1964, *35*, 647-658.

Tukey, J. Causation, regression, and path analysis. In O. Kempthorne (Ed.), *Statistics and mathematics in biology*. Ames, Iowa: Iowa State College Press, 1954.

Tubbs, J. D., Lewis, T. O., & Duran, B. S. A note on the analysis of manova model and its applications to growth curves. *Communications in Statistics*, 1975, *4*, 643-653.

Tuddenham, R. D., & Snyder, M. M. *Physical growth of boys and girls from birth to eighteen years*. Berkeley: University of California Press, 1954.

Van den Daele, L. D. Qualitative models in developmental analysis. *Developmental Psychology*, 1969, *1*, 303-310.

van't Hof, M. A., Prahl-Andersen, B., & Kowalski, C. J. A model for the study of developmental processes in dental research. *Journal of Dental Research*, 1976, *55*, 359-366.

Wald, A. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, 1940, *11*, 284-300.

Wald, A. Tests of statistical hypotheses concerning general parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 1943, *54*, 426-482.

Walker, H. M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Wallis, W. A., & Roberts, H. V. *Statistics: A new approach*. Glencoe, Ill.: Free Press, 1956.

Wampler, R. H. A report on the accuracy of some widely used least squares computer programs. *Journal of the American Statistical Association*, 1970, *65*, 549-565.

Weinback, A. P. The human growth curve II: Birth to puberty. *Growth*, 1941, *5*, 235-247.

Werner, H. The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The Concept of Development*. Minneapolis: University of Minnesota Press, 1957.

Werts, C. E., Jöreskog, K. G., & Linn, R. L. Comment on "The estimation of measurement error in panel data." *American Sociological Review*, 1971, *36*, 110-113.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. A simplex model for analysing academic growth. Educational and Psychological Measurement, 1977, 82, in press.

Wheaton, B., Muthén, B., Alwin, D. & Summers, G. Assessing reliability and stability in panel models with multiple indicators. In D. R. Heise (Ed.), Sociological Methodology, 1977, in press.

Wiley, D. E. & Harnischfeger, A. Post hoc, ergo propter hoc. Problems in the attribution of change. University of Chicago: Studies of Educative Processes, Report No. 7, 1973.

Wiley, D. E. & Hornik, R. Measurement Error and the analysis of panel data. Report #5, Studies of Educative Processes, CEMREL, August 1973.

Wiley, D. E., & Wiley, J. A. The estimation of measurement error in panel data. American Sociological Review, 1970, 35, 112-117.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wishart, J. Growth rate determinations in nutrition studies with the bacon pig, and their analysis. Biometrika, 1938, 30, 16-28.

Wishart, J. Statistical treatment of animal experiments. Journal of the Royal Statistical Society, 1939, Suppl. No. 6, 1-22.

Wohlwill, J. F. Methodology and research strategy in the study of developmental change. In L. R. Goulet and P. B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970.

Wohlwill, J. F. The age variable in psychological research. Psychological Review, 1970, 77, 49-64.

Wohlwill, J. F. The study of behavioral development. New York: Academic Press, 1973.

Wold, H. Causal inference from observational data: A review of ends and means. Journal of the Royal Statistical Society, Series A., 1956, 119, 28-61.

APPENDIX A

Book Prospectus

Nesselroade, J. R., & Baltes, P. B. (Eds.). *Longitudinal research in the behavioral sciences: Design and analysis*.

Objective

This book will present a comprehensive overview of longitudinal research methodology in the behavioral sciences (psychology, education, human development). Its special features are that (a) it will combine an easily comprehensible overview section with more technical "forward-looking" expositions, and (b) it will be co-authored by a multidisciplinary team of experts from psychology, sociology, education, and statistics. The primary substantive emphasis, however, will be on the study of behavior by means of longitudinal methodology.

Background

The volume is the product of a large-scale contract (1974-1976) which the National Institute of Education awarded to the Pennsylvania State University and on which the editors were principal investigators. The chapters have been carefully prepared and coordinated by the editors and will be (or are being) edited with a primary view on quality and substantive convergence. Expected manuscript length (typed) is 500 to 600 pages. Expected completion date of manuscripts is May 30, 1977.

Audience

There is no comparable book available in the literature. It is expected that the volume will become widely used by graduate students and researchers interested in the study of behavioral development both in the behavioral and social sciences.

There is also a rather significant likelihood that neighboring disciplines (e.g., economics, anthropology, history) will use the volume as a source manual. In general, the editors expect that the volume will enjoy a large and long-term market.

Editors

John R. Nesselroade and Paul B. Baltes
Professors of Human Development
The Pennsylvania State University
College of Human Development
University Park, PA 16802

Introduction

John R. Nesselroade and Paul B. Baltes

Part I: Longitudinal Research in Behavioral Sciences: Overview

1. History and Purpose of Longitudinal Research

John R. Nesselroade and Paul B. Baltes

2. Basic Design of Longitudinal Research

John R. Nesselroade and Paul B. Baltes

3. Analysis and Interpretation of Longitudinal Data

John R. Nesselroade and Paul B. Baltes

Part II: Advances in Longitudinal Design and Analysis

4. Components of Developmental Change: Intraindividual Change and Interindividual Differences

by Allan R. Buss
Center for Advanced Study of Theoretical Psychology
University of Alberta
Edmonton, Alberta, Canada

5. Mathematical Representations of Development Theories

by Burton Singer
Department of Mathematical Statistics
Columbia University
Broadway & West 116th Streets
New York, NY 10027

and

Seymour Spilerman
Department of Sociology
University of Wisconsin
Madison, WI 53706

6. Application of Longitudinal and Sequential Designs to the Assessment of Educational Treatments in an Age-Developmental Framework

by Larry R. Goulet
Department of Educational Psychology
University of Illinois at Urbana-Champaign
Urbana, IL 61801

7. Mathematical Description and Representation of Developmental Change Functions on the Intra- and Interindividual Levels

by Kenneth E. Guire
Center for Human Growth
and Development
University of Michigan
Ann Arbor, MI 48104

and

Charles J. Kowalski
Dental Research Institute and
Statistical Research Laboratory
University of Michigan
Ann Arbor, MI 48104

8. Assessment and Statistical Control of Subject Variables in Longitudinal Research Designs

by Paul A. Games
Department of Educational Psychology
The Pennsylvania State University
University Park, PA 16802

9. The Use of Time-Series Models in the Study of Longitudinal-Educational Research

by Carl H. Frederiksen
Chief, Research Planning and Analysis
Learning Division
National Institute on Education
1200 19th St. N.W.
Washington, D.C. 20208

and

John Rotondo
Department of Psychology
University of Virginia
Charlottesville, VA 22903

10. Univariate and Multivariate Analysis of Variance of Time-Structured Data

by R. Darrell Bock
Department of Educational Psychology
University of Chicago
Chicago, IL 60637

11. The Analysis of Categorical Data in Longitudinal Research

by J. Richard Landis
Department of Biostatistics
School of Public Health
University of Michigan
Ann Arbor, MI 48109

and

Gary G. Koch
Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27514

12. Causal Models in Longitudinal Research

by David Rogosa
Evaluation Consortium
School of Education
Stanford University
Stanford, CA 94305

13. Statistical Estimation of Structural Models in Longitudinal-Developmental Investigation

by Karl G. Jöreskog
Department of Statistics
P.O. Box 513
Uppsala University
Uppsala, Sweden

Bibliography

Author Index

Subject Index

APPENDIX B

PROJECT RELATED PAPERS

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. Cohort effects in developmental psychology: Theoretical and methodological perspectives. In A. W. Collins (Ed.), *Minnesota Symposia on Child Psychology* (Vol. 11). New York: Crowell, 1977, in press.

Baltes, P. B., & Nesselroade, J. R. *A developmentalist's view of regression toward the mean: A largely irrelevant issue to developmental research?* Unpublished manuscript, College of Human Development, The Pennsylvania State University, 1976.

Baltes, P. B., Nesselroade, J. R., & Cornelius, S. W. Environmental influence patterns and multivariate longitudinal structural change: A simulation experiment of developmental transitions. *Multivariate Behavioral Research*, submitted.

Baltes, P. B., & Schaie, K. W. On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. *American Psychologist*, 1976, *31*, 720-725.

Baltes, P. B., & Willis, S. L. Toward psychological theories of aging and development. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand-Reinhold, 1976.

Hultsch, D. F., Nesselroade, J. R., & Plemons, J. K. Learning-ability relationships in adulthood. *Human Development*, 1976, *19*, 234-247.

Nesselroade, J. R. Issues in studying developmental change in adults from a multivariate perspective. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand-Reinhold, 1976.

Nesselroade, J. R., & Baltes, P. B. Higher-order convergence of two distinct personality systems: Cattell's HSPQ and Jackson's PRF. *Multivariate Behavioral Research*, 1975, *10*, 387-408.

Schaie, K. W., & Baltes, P. B. On sequential strategies in developmental research: Description or explanation? *Human Development*, 1975, *18*, 384-390.